Proceedings of the Workshop on

Gravitation and Relativistic Astrophysics

Ahmedabad, 18-20 January 1982

Edited by

A. R. Prasanna, J. V. Narlikar & C.V. Vishveshwara

Indian Academy of Sciences/World Scientific

GRAVITATION AND
RELATIVISTIC
ASTROPHYSICS

Proceedings of the Workshop on

Gravitation and Relativistic Astrophysics

Ahmedabad, 18-20 January 1982

Edited by

A. R. Prasanna, J. V. Narlikar & C.V. Vishveshwara

Published for the

Indian Academy of Sciences
Bangalore

by

World Scientific Publishing Co.
Singapore

Indian Academy of Sciences/World Scientific

Published for the Indian Academy of Sciences by

World Scientific Publishing Co Pte Ltd.
P O Box 128
Farrer Road
Singapore 9128

ISBN 9971-966-67-0

Printed in Singapore by Richard-Clay (S. E. Asia) Pte. Ltd.

FOREWORD

Gravitation, the all-pervading interaction, brought together scientists from the leading research institutes of the country and the result of this interaction was a joint Workshop which was held at Ahmedabad, 18–20 January 1982. It is a matter of joy and satisfaction to note that this is the second occasion that the Physical Research Laboratory, the Tata Institute of Fundamental Research and the Raman Research Institute have jointly sponsored a meeting on gravitation, relativity and allied topics, the earlier occasion being the Einstein Centenary Symposium held in 1979. As in the previous occasion, the venture was supported by the Indian National Science Academy and the University Grants Commission. This Workshop held during January 1982 was the first of its kind in the country and no doubt there will be many more to follow.

The Indian Academy of Sciences kindly consented to print the proceedings of this Workshop in a book form, what is now in your hands. As the articles show, the treatment of the topics was both highly specialised and topical. An unique aspect of this Workshop is that this is the first time that astrophysicists, general relativists and field theorists have met together to discuss problems of possible mutual interest. This interaction proved rewarding, and as was emphasised by the organisers at the meeting, it is desired that this collaboration will continue and spread to many other areas of current research paving the way for a better understanding of the Universe around us.

D Lal
Physical Research Laboratory
Ahmedabad

PREFACE AND ACKNOWLEDGEMENTS

Gravitation and relativistic astrophysics are two of the most active and exciting fields of research today. The Workshop that was organised at the Physical Research Laboratory, Ahmedabad during January 18–20, 1982, concentrated on various aspects related to these subjects. The purpose of the Workshop was to offer an overall view of different topics as well as to familiarise the interested scientists with some of the details of recent developments. The participants in the Workshop were mainly experts and students in various interwoven disciplines like general relativity, field theory, astrophysics and cosmology. The proceedings of the Workshop consisted only of invited lectures divided into five sessions dealing with (1) Gravitation as Gauge Theory, (2) Gravitational Collapse, (3) Accretion Disk Dynamics, (4) Classical Cosmology and (5) Quantum Cosmology. Some of the lectures were general reviews while the others emphasised the methodology and results of specific investigations. Sufficient time was devoted to discussions after each lecture.

The present volume is a record of the lectures presented at the Workshop including some of the discussions that followed. The material has been arranged according to the same classification as was followed at the meeting. At the beginning of each section a brief outline of the material covered in the lectures is provided. It should be clear from the articles and the discussions that many unresolved problems and open issues were considered at the Workshop. Some of these still remain a challenge to physicists while new ones have inevitably been created in the meantime. The Workshop in fact marked only a beginning of fruitful discussions of such problems among scientists with common interests. We hope that more such meetings will follow in the future perhaps to consider each of the topics individually in greater depth and detail.

It is a great pleasure to thank the sponsoring institutes—the Physical Research Laboratory, Ahmedabad, the Raman Research Institute, Bangalore, and the Tata Institute of Fundamental Research, Bombay as well as the Indian National Science Academy and the University Grants Commission for their generous grants which made the Workshop possible.

We are indebted to all the speakers and participants and in particular to Professor P. C. Vaidya for writing the introduction to the section on classical cosmology. We wish to express our sincere appreciation for the valuable help rendered by many of our colleagues at various stages in both scientific and administrative matters.

We are grateful to the Indian Academy of Sciences for undertaking the publication of this volume.

A. R. Prasanna
J. V. Narlikar
C. V. Vishveshwara

CONTENTS

§ I. GRAVITATION AS A GAUGE THEORY

INTRODUCTION

Gravitation, the most fundamental of all the forces in Nature, governs the entire Universe through its manifestation as curvature of space-time. As a description of the macrocosom though gravitation has been perfectly well described through the geometry of the manifold, it has still not been possible to get a proper description of gravitation in the microworld. However attempts have been made at various stages to get a theory of Quantum Gravity but none could be said as in the right direction except possibly the gauge group description of gravitation. If one considers the other interactions in Nature, it is now well established that the electroweak interaction is a gauge theory and the unification of electroweak with strong interaction also is a gauge theory. With this in background if one looks for a possible unification of all the four interactions the basic structure that underlies could be a gauge group. Hence it is not just fortutitous that gravitation by itself is a gauge theory with the Poincare group acting as the gauge group. But the only hurdle in this has been the physical understanding of the role of the invariants associated with the Lie algebra of this group, when one considers the general covariance. As far as the Poincare group is concerned, if it is the complete symmetry group of the manifold in the sense of having a global inertial frame then it is very clear that the invariants associated with its Lie algebra represent mass and spin of the elementary particle. But then we do not have gravitation in that manifold. Once we introduce gravitation and the general covariance treating Poincare group as the local gauge group one will now have in principle, curvature and torsion of the manifold both non-zero. Whereas curvature represents the energy-momentum of the system, it is not clear as to the role of torsion. If one takes in the point of view that torsion represents the spin-spin contact interaction then in the framework of general relativity one would get this interaction in the Lagrangian with a coupling constant κ^2 wherein $\kappa \sim 10^{-39}$, and thus completely negligible. Apart from this it has also been pointed out that torsion being not determined by the metric does not represent any new topological invariant of the manifold and hence may not be significant at all.

With this in the background it is obvious that one should consider in detail the aspects of gravity as a gauge theory and get a complete understanding of the associated elements of the space-time manifold. In this papers and discussion that follows, some of the basic features of gravity as a gauge theory have been presented. It begins with a masterly introduction to the gauge approach to classical gravity *a la* Utiyama and Kibble, followed by a new attempt to understand long range spin-spin interaction. Finally a lucid introduction to supergravity—a theory attempting to treat gravitation both at classical and quantum level—is presented.

It is clear that there are a number of unsolved problems in this fertile field as is borne out in the discussions.

Gauge approach to classical gravity

N MUKUNDA
Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560012

1. Introduction

The purpose of this paper is to give a pedagogical introduction to the treatment of gravitation theory as a gauge theory, covering essentially the two early and well-known papers of Utiyama (1956) and Kibble (1961). The basic idea is to apply to the Poincaré group the same kind of gauge principle that leads, in the case of a non-Abelian internal symmetry, to the well-known Yang–Mills theory. One starts with a Lagrangian density L_M describing the dynamics of a set of matter fields ϕ, and possessing invariance either under an internal symmetry group G or the Poincaré group $\mathscr{P}$ as the case may be. One then alters L_M in a minimal way, with the introduction of suitable new fields, to achieve invariance under the gauged versions G', $\mathscr{P}'$ of G, $\mathscr{P}$ corresponding to making the parameters of G, $\mathscr{P}$ arbitrary functions of space-time. Finally one examines possible forms for the free Lagrangian for the new fields introduced in this process, and analyses the structure and consistency of the new field equations.

2. Gauging an internal symmetry group—Yang–Mills theory

Consider a multiplet of matter fields $\phi(x)$ belonging to some representation of a semi-simple compact Lie group G; and whose dynamics, described by a matter Lagrangian density $L_M(\phi; \partial_\mu\phi)$, is invariant under G by virtue of the invariance of L_M itself. Let the action of an infinitesimal element of G on $\phi(x)$ be given by the equations

$$\delta\phi(x) = \varepsilon^a T_a \phi(x). \tag{1}$$

Here ε^a, $a = 1, 2, \ldots, n$ are the parameters of the infinitesimal transformation and T_a are the generator matrices for the representation of G to which ϕ belongs. For the space-time derivatives of $\phi(x)$ one has the transformation law

$$\delta\partial_\mu\phi(x) = \varepsilon^a T_a \partial_\mu \phi(x). \tag{2}$$

The invariance of $L_M(\phi; \partial\phi)$ under (1) and (2) implies n identities characterising the functional form of L_M; they can be stated as

$$\delta L_M(\phi; \partial\phi) = 0 \Rightarrow$$

$$\frac{\partial L_M}{\partial\phi} T_a\phi + \frac{\partial L_M}{\partial\phi_\mu} T_a\phi_\mu = 0,\ L_M = L_M(\phi; \phi_\mu), \tag{3}$$

where in the last line the arguments ϕ_μ need not be the space-time derivatives of ϕ but can be anything at all.

The transformation (1) corresponding to an internal symmetry is very much like a point transformation in mechanics in that the change in ϕ involves ϕ alone, not the derivatives of ϕ.

To gauge the group G and arrive at an extended group G' is to make the parameters ε^a arbitrary independent functions of space-time. While the change in ϕ under G' is still given by (1), the change in $\partial\phi$ now has an extra piece when compared to (2):

$$\delta\partial_\mu\phi(x) = \varepsilon^a T_a \partial_\mu\phi(x) + \partial_\mu\varepsilon^a \cdot T_a\phi(x). \tag{4}$$

This extra piece, however, involves ϕ and not $\partial\phi$. The original Lagrangian density $L_M(\phi; \partial\phi)$ is naturally not expected to be invariant under G'; indeed we find

$$\delta L_M(\phi; \partial\phi) = J^\mu_a \partial_\mu \varepsilon^a,$$
$$J^\mu_a = \frac{\partial L_M(\phi; \partial\phi)}{\partial\partial_\mu\phi} T_a\phi. \tag{5}$$

The n currents J^μ_a are conserved in the G-invariant theory.

What must be done to get a new matter Lagrangian invariant under G'? Recognising that in the identities (3), ϕ_μ need not be $\partial_\mu\phi$, and that in (4) the extra piece is linear in ϕ, we can see: if we can construct a modified derivative $D_\mu\phi$, differing from $\partial_\mu\phi$ by terms linear in ϕ, and transforming under G' *via*

$$\delta D_\mu\phi = \varepsilon^a T_a D_\mu\phi, \tag{6}$$

then $L_M(\phi; D\phi)$ will be G'-invariant. The way to do this is to introduce $4n$ new fields $A^a_\mu(x)$—the Yang–Mills gauge potentials—and to define

$$D_\mu\phi = (\partial_\mu + A^a_\mu T_a)\phi. \tag{7}$$

Then we will secure (6) provided the change of A^a_μ under G' reads

$$\delta A^a_\mu = \varepsilon^b f^a_{bc} A^c_\mu - \partial_\mu\varepsilon^a, \tag{8}$$

where the structure constants f^a_{bc} of G are identified by

$$[T_a, T_b] = f^c_{ab} T_c. \tag{9}$$

The potentials A^a_μ transform in linear inhomogeneous fashion under G' and in that sense do not belong to a representation of G', unlike ϕ. The infinitesimal transformation laws of ϕ and A^a_μ under G' can be integrated to finite forms, but we omit the details. The operator D_μ of covariant differentiation can be applied to any field belonging to any linear representation of G', and the result will belong to the same representation. In that sense it cannot be applied to A^a_μ without further qualification.

In order to build a G'-invariant free Lagrangian density for the new field A^a_μ, we first evaluate the commutator of two D's:

$$[D_\mu, D_\nu] = F^a_{\nu\mu} T_a,$$
$$F^a_{\nu\mu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - f^a_{bc} A^b_\nu A^c_\mu. \tag{10}$$

Unlike the potential A, this field strength transforms in a linear and homogeneous way under G':

$$\delta F^a_{\mu\nu} = \varepsilon^b f^a_{bc} F^c_{\mu\nu}. \tag{11}$$

Thus it belongs to the adjoint representation of G, with generator matrices made up from the structure constants themselves.

Let us look for a free Lagrangian density $L_0(A, \partial A)$ for the gauge field, depending only on A and its first space-time derivatives. The requirement that it be invariant under G', *i.e.* when A undergoes the transformation (8), leads first to the result that L_0 must be a function exclusively of $F^a_{\mu\nu}$, so we have to deal with $L_0(F)$. Now for $L_0(F)$ the requirements of G' invariance and G invariance essentially coincide, because of (11), and they imply:

$$G^{\mu\nu}_a = \frac{\partial L_0(F)}{\partial F^a_{\mu\nu}}, \quad G^{\mu\nu}_a f^a_{bc} F^c_{\mu\nu} = 0. \tag{12}$$

The total Lagrangian density possessing G' invariance is

$$L_I = L_M(\phi; D\phi) + L_0(F), \tag{13}$$

it is built from a matter Lagrangian whose functional form is restricted by the n conditions (3), and from a gauge-field Lagrangian restricted by the n conditions (12).

Usually one assumes the background space-time to be flat and Minkowskian, and then the simplest choice for $L_0(F)$ obeying (12) is

$$L_0(F) = \tfrac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a, \quad G^{\mu\nu}_a = F^{\mu\nu}_a. \tag{14}$$

The raising and lowering of the indices a, b of the adjoint representation of G is done in the usual manner.

The Euler–Lagrange field equations corresponding to the new fields A^a_μ are:

$$\frac{\delta L_I}{\delta A^a_\mu} = 0 \Rightarrow D_\nu G^{\nu\mu}_a = \frac{\partial L_M}{\partial A^a_\mu} = \frac{\partial L_M}{\partial D_\mu \phi} T_a \phi = J^\mu_a, \tag{15}$$

where the matter current J^μ_a here differs from what appeared in (5) by the replacement $\partial\phi \rightarrow D\phi$. Now the G' invariance of L_0 and of the associated action implies that the object appearing on the left side of the field equation (15) obeys an identity:

$$D_\mu D_\nu G^{\nu\mu}_a = 0. \tag{16}$$

This (Bianchi) identity does not depend for its validity on any field equations being obeyed. Consistency of (15) demands that the matter current J^μ_a must obey a corresponding relation. One can now show that the G'-invariance of $L_M(\phi; D\phi)$ leads in turn to another identity:

$$D_\mu J^\mu_a = \frac{\delta L_M}{\delta\phi} T_a \phi. \tag{17}$$

Thus we are assured of the consistency of the A-field equations precisely when the matter fields obey **their** equations of motion.

3. Gauging the Poincaré group

A gauged internal symmetry group G' acts only on the matter and gauge fields, not on the space-time coordinates. We now see how the method of the previous section must be extended to gauge the Poincaré group which does have an effect on space-time itself.

We begin with a set of matter fields $\phi(x)$ defined on a flat space-time background with

Cartesian coordinates x^μ. The Poincaré invariant matter Lagrangian density is $L_M(\phi;\partial\phi)$. The geometrical action of an infinitesimal Poincaré transformation, on x^μ and $\phi(x)$, is given by the system of equations

$$x'^\mu \simeq x^\mu + \varepsilon^{\mu\nu}x_\nu + a^\mu,$$

$$\phi'(x') \simeq \phi(x) + \tfrac{1}{2}\varepsilon^{\alpha\beta}S_{\alpha\beta}\phi(x). \tag{18}$$

The parameters a^μ specify an infinitesimal translation; and the six $\varepsilon^{\mu\nu} = -\varepsilon^{\nu\mu}$ an infinitesimal homogeneous Lorentz rotation. The $S_{\alpha\beta}$ are the generators of the representation of the Lorentz group (more precisely $SL(2, C)$) to which ϕ belongs. We shall find it more convenient to 'undo' the transformation on x^μ in (18) and deal simply with the change in functional form of ϕ induced by an infinitesimal element of $\mathscr{P}$. Thus we write

$$\delta\phi(x) = \tfrac{1}{2}\varepsilon^{\alpha\beta}S_{\alpha\beta}\phi(x) - \xi^\nu\partial_\nu\phi(x),$$

$$\xi^\nu = \varepsilon^{\nu\lambda}x_\lambda + a^\nu. \tag{19}$$

Already we see a qualitative difference between (19) and (1): in the former the change in ϕ involves $\partial\phi$ as well as ϕ. Thus (19) is like a dynamical, velocity-dependent transformation in mechanics, rather than a point transformation. The effect of $\mathscr{P}$ on $\partial\phi$ follows from (19):

$$\delta\partial_\mu\phi(x) = \tfrac{1}{2}\varepsilon^{\alpha\beta}S_{\alpha\beta}\partial_\mu\phi(x) + \varepsilon_\mu{}^\nu\partial_\nu\phi(x) - \xi^\nu\partial_\nu\partial_\mu\phi(x). \tag{20}$$

The geometrical invariance of the action based on $L_M(\phi;\partial\phi)$ under a Poincaré transformation can be stated as the property

$$\delta L_M(\phi;\partial\phi) = \partial_\mu(-(\varepsilon^{\mu\nu}x_\nu + a^\mu)L_M), \tag{21}$$

when $\delta\phi$ and $\delta\partial_\mu\phi$ are given by (19) and (20) respectively. On analysing (21), we find that it says two things: (i) there should be no explicit x dependence in L_M (and this has any way been assumed); (ii) the functional form of L_M is such that the six identities

$$\frac{\partial L_M}{\partial\phi}S_{\alpha\beta}\phi + \frac{\partial L_M}{\partial\phi_\mu}(S_{\alpha\beta}\phi_\mu + \eta_{\mu\alpha}\phi_\beta - \eta_{\mu\beta}\phi_\alpha) = 0,$$

$$L_M = L_M(\phi;\phi_\mu), \tag{22}$$

hold, with the variables ϕ_μ being anything at all (compare (20)).

Let us now gauge the group $\mathscr{P}$ and arrive at a group $\mathscr{P}'$ by making the parameters $\varepsilon^{\mu\nu}$, a^μ in (19) arbitrary independent functions of x. When this is done, we no longer maintain that the expression ξ^ν in (19) has an x-independent part and a part linear in x with $\varepsilon^{\nu\lambda}$ for coefficient—making $\varepsilon^{\mu\nu}$ and a^μ arbitrary independent functions of x is entirely equivalent to regarding $\varepsilon^{\mu\nu}$ and ξ^μ as independent functions of x. By the same token, we have no longer any connection between the parameters of the infinitesimal $SL(2, C)$ rotation represented by the first term in $\delta\phi$ in (19), and the second term in $\delta\phi$—these two pieces of the transformation law of ϕ under $\mathscr{P}'$ are completely dissociated. For this reason we shall hereafter use Latin superscripts and subscripts as vector indices for ε and the $SL(2, C)$ generators S, while Greek indices will appear on ξ, on x and on partial derivatives with respect to x. With this understanding, an infinitesimal element of $\mathscr{P}'$ is specified by ten independent infinitesimal functions $\varepsilon^{ij}(x)$, $\xi^\nu(x)$ and its action on ϕ is:

$$\delta\phi(x) = \tfrac{1}{2}\varepsilon^{ij}S_{ij}\phi(x) - \xi^\nu\partial_\nu\phi(x). \tag{23}$$

The accompanying law for $\partial\phi$ reads:

$$\delta\partial_\mu\phi(x) = \tfrac{1}{2}\varepsilon^{ij}S_{ij}\partial_\mu\phi(x) - \partial_\mu\xi^\nu\cdot\partial_\nu\phi - \xi^\nu\partial_\nu\partial_\mu\phi + \tfrac{1}{2}\partial_\mu\varepsilon^{ij}\cdot S_{ij}\phi. \tag{24}$$

The second term in (24) replaces the second term in (20), while the fourth term in (24) is new and is analogous to the second term in (4) in the sense of being linear in ϕ.

The original Poincaré invariant Lagrangian density $L_M(\phi;\partial\phi)$ is of course not expected to be invariant (even up to a four-divergence) under the action of $\mathscr{P}'$. Indeed we find:

$$\delta L_M(\phi;\partial\phi) = \partial_\mu(-\xi^\mu L_M) + \tfrac{1}{2}M^\mu_{ij}\partial_\mu\varepsilon^{ij} - T^\mu_\nu\partial_\mu a^\nu, \tag{25}$$

where M^μ_{ij} and T^μ_ν are the angular momentum and energy momentum densities conserved in the $\mathscr{P}$-invariant theory (compare with (5)).

We must now see how to modify $L_M(\phi,\partial\phi)$ in a natural way to achieve $\mathscr{P}'$-invariance. Following the method of the previous section, we might try to define a new derivative of ϕ whose transformation law under $\mathscr{P}'$ resembles that of $\partial\phi$ under $\mathscr{P}$; and we might expect that the problem will be solved by substituting this new derivative for $\partial\phi$ in L_M since in the identities (22) there is no requirement that ϕ_μ be necessarily $\partial_\mu\phi$. Comparing (24) and (20), and recalling the introduction of Latin indices on ε^{ij}, we are led to search for a generalised derivative ϕ_k of ϕ whose change under $\mathscr{P}'$ shall read

$$\delta\phi_k = \tfrac{1}{2}\varepsilon^{ij}S_{ij}\phi_k + \varepsilon_k{}^l\phi_l - \xi^\nu\partial_\nu\phi_k. \tag{26}$$

If we succeed in constructing such a ϕ_k, we then find that $L_M(\phi;\phi_k)$ is almost but not quite invariant (upto a divergence) under $\mathscr{P}'$: the identities (22) obeyed by L_M only give us

$$\delta L_M(\phi;\phi_k) = -\xi^\nu\partial_\nu L_M(\phi;\phi_k). \tag{27}$$

Clearly one more step, not needed in the previous section, is needed here to achieve $\mathscr{P}'$ invariance of the action of matter: we need a multiplier Δ, say, with a suitable $\mathscr{P}'$ transformation law so that $\mathscr{L}_M = \Delta L_M$ is indeed $\mathscr{P}'$ invariant upto a divergence:

$$\mathscr{L}_M = \Delta L_M(\phi;\phi_k),$$
$$\delta\Delta = \partial_\mu(-\xi^\mu\Delta),\ \delta\mathscr{L}_M = \partial_\mu(-\xi^\mu\mathscr{L}_M). \tag{28}$$

This multiplier Δ will of course have to be formed out of the new fields that will appear in the construction of ϕ_k out of $\partial_\mu\phi$.

In the rest of this section, we describe the construction of ϕ_k and Δ, so that the procedure for gauging $\mathscr{P}$ as far as the matter fields are concerned will be clear.

The passage from $\partial_\mu\phi$ transforming *via* (24) to ϕ_k transforming *via* (26) will be effected in two stages. First we define an auxiliary object $\phi_{|\mu}$ in whose law of change under $\mathscr{P}'$ no term like the last one in (24) appears—by a procedure exactly like the one in the previous section, we introduce a gauge potential A^{ij}_μ with respect to $SL(2,C)$ to absorb the $\partial_\mu\varepsilon^{ij}$ term in $\delta\partial_\mu\phi$. Thus with the definition

$$\phi_{|\mu} = (\partial_\mu + \tfrac{1}{2}A^{ij}_\mu S_{ij})\phi, \tag{29}$$

we will attain the transformation law

$$\delta\phi_{|\mu} = \tfrac{1}{2}\varepsilon^{ij}S_{ij}\phi_{|\mu} - \partial_\mu\xi^\nu\ \phi_{|\nu} - \xi^\nu\partial_\nu\phi_{|\mu} \tag{30}$$

provided A^{ij}_μ transforms under $\mathscr{P}'$ as follows:

$$\delta A^{ij}_\mu = \varepsilon^i{}_k A^{kj}_\mu + \varepsilon^j{}_k A^{ik}_\mu - \partial_\mu \xi^\nu \cdot A^{ij}_\nu - \xi^\nu \partial_\nu A^{ij}_\mu - \partial_\mu \varepsilon^{ij}. \tag{31}$$

This first step to go from $\partial_\mu \phi$ to $\phi_{|\mu}$ has thus necessitated the introduction of a 24-component $SL(2, C)$ gauge potential. At the next step, we must make up ϕ_k out of $\phi_{|\mu}$ so that the second term in (30) gives way to the second term in (26). This passage from $\phi_{|\mu}$ to ϕ_k is qualitatively different from the earlier passage from $\partial_\mu \phi$ to $\phi_{|\mu}$: in the latter case, exactly as with an internal symmetry, the extra piece was linear in ϕ, whereas in the former case ϕ_k will have to be linear homogeneous in $\phi_{|\mu}$. This thus calls for 16 new fields $h_k{}^\mu$ which can be viewed as a 4×4 matrix, so that we may write

$$\phi_k = h_k{}^\mu \phi_{|\mu} = h_k{}^\mu (\partial_\mu + \tfrac{1}{2} A^{ij}_\mu S_{ij}) \phi. \tag{32}$$

Then (26) will be guaranteed provided under $\mathscr{P}'$

$$\delta h_k{}^\mu = \varepsilon_k{}^l h_l{}^\mu + \partial_\nu \xi^\mu . h_k{}^\nu - \xi^\nu \partial_\nu h_k{}^\mu. \tag{33}$$

The above steps can probably be more easily grasped if we also give the effect of a finite element of $\mathscr{P}'$ on ϕ, A and h. A finite element of the group $\mathscr{P}'$ is specified in this way: we have a general coordinate transformation $x^\mu \to x'^\mu(x)$ in which the space-time labels assigned to each point may change; and we have an element $\lambda(x) \varepsilon SL(2, C)$ at each point with (original) coordinates x. Thus an element of $\mathscr{P}'$ is denoted by $[x'(x); \Lambda(x)]$. The finite form of (23) is

$$\phi'(x') = D[\Lambda(x)] \phi(x), \tag{34}$$

where D is a matrix in the representation of $SL(2, C)$ to which ϕ belongs in the original Poincaré invariant theory. From (34) it is particularly clear that in gauging $\mathscr{P}$ to arrive at $\mathscr{P}'$, the local $SL(2, C)$ rotations have been completely dissociated from general coordinate transformations (GCT). Under a pure GCT, each component of ϕ is now treated as a scalar field. Any spinor, vector or tensor behaviour that ϕ has in the special relativistic theory is now assigned to be its behaviour under local $SL(2, C)$ rotations. With ϕ transforming as above, its gradient follows the law

$$\partial'_\mu \phi'(x') = \frac{\partial x^\nu}{\partial x'^\mu} D[\Lambda(x)] \{\partial_\nu \phi(x) + D[\Lambda(x)]^{-1} \partial_\nu D[\Lambda(x)] \cdot \phi(x)\}. \tag{35}$$

The finite forms of (30) and (31) are respectively

$$\phi'_{|\mu}(x') = \frac{\partial x^\nu}{\partial x'^\mu} D[\Lambda(x)] \phi_{|\nu}(x), \tag{36a}$$

$$A'^{ij}_\mu(x') = \frac{\partial x^\nu}{\partial x'^\mu} [\Lambda^i{}_k(x) \Lambda^j{}_l(x) A^{kl}_\nu(x) - \partial_\nu \Lambda^i{}_k(x) \cdot \Lambda^{jk}(x)]. \tag{36b}$$

[We have for simplicity not differentiated between Λ as an element of $SL(2, C)$ in (34), and as a 4×4 Lorentz matrix in (36)].

The finite forms of (26) and (33) are respectively

$$\phi'_k(x') = \Lambda_k{}^j(x) D[\Lambda(x)] \phi_j(x), \tag{37a}$$

$$h'^\mu_k(x') = \frac{\partial x'^\mu}{\partial x^\nu} \Lambda_k{}^j(x) h_j{}^\nu(x), \tag{37b}$$

It is now particularly easy to see that A and h play qualitatively different roles. A^{ij}_{μ} is a true gauge potential with an inhomogeneous transformation law under local $SL(2, C)$ rotations, designed to compensate for the second term within the parenthesis on the right side of (35). On the other hand, the h field converts the covariant vector behaviour of $\phi_{|\mu}$ under GCT to the local $SL(2, C)$ vector behaviour of ϕ_k: compare (36a) with (37a).

The finite form of the transformation law for the multiplier Δ, (28), is

$$\Delta'(x') = \left|\frac{\partial x}{\partial x'}\right| \Delta(x), \tag{38}$$

where the Jacobian of the GCT comes in. Comparing with (37), we have the natural choice

$$\Delta(x) = \{\det [h_k{}^{\mu}(x)]\}^{-1}. \tag{39}$$

We must assume of course that $(h_k{}^{\mu})$ is a nonsingular matrix. The way to gauge the matter Lagrangian density to get $\mathscr{P}'$ invariance is now clear. We introduce 40 new variables A^{ij}_{μ}, $h_k{}^{\mu}$, set up the generalised derivative ϕ_k of (32) and make the transition

$$\mathscr{P}\text{-invariant } L_M(\phi; \partial\phi) \rightarrow$$

$$\mathscr{P}'\text{-invariant } \mathscr{L}_M = \Delta(h) L_M(\phi; \phi_k). \tag{40}$$

(of course, we mean here that the corresponding actions have the stated invariance properties).

We conclude this section with the remark that in checking the consistency of the finite and infinitesimal changes of various objects under $\mathscr{P}'$, the form of an infinitesimal element of $\mathscr{P}'$ is to be taken as

$$[x'(x); \Lambda^k{}_j(x)] = [x^{\mu} + \xi^{\mu}(x); \delta^k_j + \varepsilon^k{}_j(x)]. \tag{41}$$

4. Independent gravitational variables

While the new fields A^{ij}_{μ}, $h_k{}^{\mu}$ (40 in all) must transform under $\mathscr{P}'$ according to (31), (33) (equally well (36b), (37b)), their kinematical or dynamical independence is at this point an open question. After examining the possibilities allowed, we can go on to build the free Lagrangian for these new fields.

Denote by D_{μ} the generalised derivative in (29) as in (7):

$$D_{\mu} = \partial_{\mu} + \tfrac{1}{2} A^{ij}_{\mu} S_{ij}. \tag{42}$$

Its property is that if a (matter) field ϕ is a GCT scalar but belongs to some $SL(2, C)$ representation with generators S_{ij}, then $D\phi$ is a covariant GCT vector but retains the $SL(2, C)$ representation of ϕ. Denote this by:

$$\phi \in \{SL(2, C) \text{ rep.}\} \otimes \{\text{GCT scalar}\} \Rightarrow$$

$$D\phi \in \{\text{same } SL(2, C) \text{ rep.}\} \otimes \{\text{GCT cov't. vector}\}. \tag{43}$$

If ϕ is a field belonging to a nontrivial tensor representation of GCT and we wish its generalised derivative to behave reasonably, we must construct an object ∇_{μ} by adding an extra piece to D_{μ} of (42): this piece will involve the generators X^{γ}_{β} of the tensor representation of $GL(4, R)$ appropriate to the tensor nature of ϕ. We then write

$$\nabla_{\mu} = \partial_{\mu} + \tfrac{1}{2} A^{ij}_{\mu} S_{ij} + \Gamma^{\beta}_{\mu\gamma} X^{\gamma}_{\beta}. \tag{44}$$

and demand of ∇ that

$$\phi \,\varepsilon \{SL(2,C)\,\text{rep.}\} \otimes \{\text{GCT tensor}\} \Rightarrow$$
$$\nabla\phi \,\varepsilon \{\text{same } SL(2,C) \text{ rep.}\} \otimes \{\text{same GCT tensor} \otimes \text{GCT cov't. vector}\}. \quad (45)$$

This requirement on ∇_μ will be met if A^{ij}_μ transforms as before while Γ transforms as an affine connection under GCT and is an $SL(2,C)$ scalar:

$$\Gamma'^{\beta}_{\mu\alpha}(x') = \frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\alpha}\frac{\partial x'^\beta}{\partial x^\rho}\Gamma^\rho_{\lambda\sigma}(x) + \frac{\partial x'^\beta}{\partial x^\rho}\frac{\partial^2 x^\rho}{\partial x'^\mu \partial x'^\alpha}. \quad (46)$$

For matter fields of course ∇_μ and D_μ coincide. Conversely, if ϕ is an $SL(2,C)$ scalar, ∇_μ and δ_μ coincide, where

$$\delta_\mu = \partial_\mu + \Gamma^\beta_{\mu\alpha}X^\alpha_\beta. \quad (47)$$

One can also say: ∇_μ is applicable only to fields with well-defined $SL(2,C)$ as well as GCT representation properties; D_μ can act on any field with definite $SL(2,C)$ representation character, even if it has no definite GCT behaviour; and δ_μ can be applied to any GCT tensor. In all these statements, one can view A^{ij}_μ and $\Gamma^\beta_{\mu\alpha}$ as independent entities, each with a specific behaviour under $\mathscr{P}'$.

On the same basis, one can also compute the commutator among the ∇_μ:

$$[\nabla_\mu, \nabla_\nu] = \tfrac{1}{2}F^{ij}_{\mu\nu}S_{ij} + R^\beta_{\alpha\mu\nu}X^\alpha_\beta + T^\lambda_{\mu\nu}\nabla_\lambda,$$
$$F^{ij}_{\mu\nu} = \partial_\nu A^{ij}_\mu - \partial_\mu A^{ij}_\nu + A^{il}_\mu A^{j}{}_{l\nu} - A^{il}_\nu A^{j}{}_{l\mu},$$
$$R^\beta_{\alpha\mu\nu} = \partial_\mu \Gamma^\beta_{\nu\alpha} - \partial_\nu \Gamma^\beta_{\mu\alpha} + \Gamma^\beta_{\mu\sigma}\Gamma^\sigma_{\nu\alpha} - \Gamma^\beta_{\nu\sigma}\Gamma^\sigma_{\mu\alpha},$$
$$T^\lambda_{\mu\nu} = \Gamma^\lambda_{\mu\nu} - \Gamma^\lambda_{\nu\mu}. \quad (48)$$

F here is the $SL(2,C)$ gauge field strength, exactly analogous to the situation in §2; R is the curvature; and T the torsion. They are all homogeneously transforming quantities under $\mathscr{P}'$: F is an antisymmetric tensor under $SL(2,C)$ and also under GCT (covariant); R and T are $SL(2,C)$ scalars and have obvious GCT behaviours.

The field $h_k{}^\mu$ is a vector under $SL(2,C)$ and a contravariant vector under GCT. It is interpreted as a Vierbein system. The inverse matrix is denoted as $b^k{}_\mu$:

$$b^k{}_\mu h_k{}^\mu = \delta^\nu_\mu. \quad (49)$$

It is again an $SL(2,C)$ vector and a GCT covariant vector. The metric tensor is defined through

$$g_{\mu\nu} = b^k{}_\mu b_{k\nu}, \quad (50)$$

where Latin indices are raised and lowered using the Minkowski metric. We now use the h-field to connect the spaces of $SL(2,C)$ tensors and of GCT tensors by postulating

$$\nabla_\lambda h_k{}^\mu = 0. \quad (51)$$

This system of equations helps us trace the relationships among the 24 A's, the 16 h's and the 64 Γ's. Its meaning of course is that application of ∇ commutes with transferring $SL(2,C)$ tensor behaviour to corresponding GCT tensor behaviour and vice versa with the help of h and b.

Suitably identifying S_{ij} and X^α_β so as to apply to $h_k{}^\mu$, we find that there are two ways of presenting the content of (51). *Viewpoint* (*a*) is to regard (51) as determining Γ in terms of

h, A and the derivatives of h (or equally well of b):

$$\Gamma^{\beta}_{\mu\alpha} = h_k{}^{\beta}(\partial_\mu b^k{}_\alpha + A^{kl}_\mu b_{l\alpha}) = h_k{}^{\beta} D_\mu b^k{}_\alpha. \tag{52}$$

Naturally this is consistent in the sense that the stated $\mathscr{P}'$ transformation laws of h and A ensure the one desired for Γ. We may thus view the 40 A's and h's as independent fields (unless we impose more conditions on them), and the 64 Γ's as dependent fields.

Viewpoint (*b*) is to regard A^{ij}_μ as a derived object, formed from Γ and h. For this, we express (51) as

$$A^{ij}_\mu = h^{i\nu}(\partial_\mu b^j{}_\nu - \Gamma^{\rho}_{\mu\nu} b^j{}_\rho) = h^{i\nu}\delta_\mu b^j{}_\nu. \tag{53}$$

The stated $\mathscr{P}'$ transformation laws of Γ and h do lead, *via* this identification, to the proper law for A. However it would be incorrect to conclude that the 64 components of Γ and the 16 of h are all independent! One must insist that the expression in (53) for A^{ij}_μ is antisymmetric in i and j. This requirement, it turns out, is tantamount to recognising that, on account of (51), both $b^k{}_\mu$ and the metric $g_{\mu\nu}$ are covariantly constant; and the latter fact does not involve A^{ij}_μ at all! In other words,

$$\nabla_\mu h = 0 \Rightarrow \nabla_\mu g = 0 \Rightarrow \delta_\mu g = 0 \Rightarrow$$
$$\Gamma^{\beta}_{\mu\alpha} = \Gamma^{(0)\beta}_{\mu\alpha} + B^{\beta}{}_{\mu\alpha},\ B_{\beta\mu\alpha} + B_{\alpha\mu\beta} = 0. \tag{54}$$

Here $\Gamma^{(0)}$ is the (symmetric) Christoffel connection determined by $g_{\mu\nu}$ (and so ultimately by h), and B is a third rank tensor. So in this second viewpoint the independent fields are the 16 h's and the 24 B's, adding up again to 40 fields exactly as in the first viewpoint. The tensor B is essentially the same as the torsion:

$$T^{\beta}_{\mu\alpha} = B^{\beta}{}_{\mu\alpha} - B^{\beta}{}_{\alpha\mu},$$
$$B_{\beta\mu\alpha} = \tfrac{1}{2}(T_{\beta\mu\alpha} - T_{\alpha\mu\beta} - T_{\mu\alpha\beta}). \tag{55}$$

Thus we see that the number of independent gravitational field variables is 40 except for any further restrictions we may impose; these 40 may be taken to be the Vierbein field h and the $SL(2, C)$ gauge potential A; or we may take them to be the Vierbein h and the torsion tensor T. Without implying that the vanishing of the one implies the vanishing of the other, we can say that A^{ij}_μ and $T^{\lambda}_{\mu\nu}$ are two aspects of the same thing or that they represent the same degrees of freedom. Expression of one in terms of the other brings in h and its derivatives:

$$T^{\beta}_{\mu\alpha} = A^{\beta}_{\alpha\mu} - A^{\beta}_{\mu\alpha} + h_k{}^{\beta}(\partial_\mu b^k{}_\alpha - \partial_\alpha b^k{}_\mu),$$
$$A^{\beta}_{\alpha\mu} = h_i{}^{\beta} b_{j\alpha} A^{ij}_\mu; \tag{56a}$$

$$A_{\alpha\beta\mu} = \tfrac{1}{2}(T_{\alpha\mu\beta} - T_{\beta\mu\alpha} + T_{\mu\alpha\beta})$$
$$+ \tfrac{1}{2}(b_{k\alpha}(\partial_\beta b^k{}_\mu - \partial_\mu b^k{}_\beta) + b_{k\mu}\partial_\beta b^k{}_\alpha - (\alpha \leftrightarrow \beta)). \tag{56b}$$

We get another useful consequence of the postulate (51) by applying both sides of (48) to the h field. The result is:

$$F^{ij}_{\mu\nu} = b^i{}_\beta h^{j\alpha} R^{\beta}_{\alpha\mu\nu} \tag{57}$$

Thus the number of true tensor quantities available for constructing the free gravitational Lagrangian is three—F or equivalently R, the torsion T and the Vierbein h

5. Gravitational Lagrangian and field equations

There are two options available here, depending on the number of independent field components admitted before constructing the Lagrangian.

Option I This is essentially based on viewpoint (b) of the previous section. It is the method given in the main body of Utiyama's paper and described in Weinberg's book. It can be viewed as a minimal extension of the metric theory of gravitation needed to accommodate spinor matter fields.

The torsion tensor T is assumed to be zero. Then the affine connection Γ is symmetric and equals the Christoffel connection $\Gamma^{(0)}$ determined by $g_{\mu\nu}$ and hence by $h_k{}^\mu$. The curvature tensor is equal to the Riemann–Christoffel tensor determined by g and its derivatives. The independent gravitational field variables are the 16 Vierbein components $h_k{}^\mu$—physically, ten metric components plus six "$SL(2, C)$ parameters" at each point x. The $SL(2, C)$ gauge potential A^{ij}_μ is a derived object (as is $g_{\mu\nu}$):

$$A^{ij}_\mu = h^{i\nu}(\partial_\mu b^j{}_\nu - \Gamma^{(0)\rho}_{\mu\nu} b^j{}_\rho). \tag{58}$$

The simplest choice for the gravitational Lagrangian density is the standard one in the metric theory leading to the Hilbert action:

$$\mathscr{L}_G = \tfrac{1}{2}\Delta R = \tfrac{1}{2}(-\det(g_{\mu\nu}))^{1/2} R(g, \partial g, \partial\partial g), \tag{59}$$

where R is the curvature scalar. Thus the total Lagrangian density is

$$\begin{aligned}\mathscr{L}_I &= \mathscr{L}_M + \mathscr{L}_G \\ &= \Delta(h)[L_M(\phi; \phi_k) + \tfrac{1}{2}R(g, \partial g, \partial\partial g)].\end{aligned} \tag{60}$$

In this treatment, the independent fields are h and ϕ; A^{ij}_μ is not expressible in terms of $g_{\mu\nu}$ alone, but the affine connection Γ and the curvature tensor R depend on $g_{\mu\nu}$ alone.

To examine the consistency of the gravitational field equations, we need the Eulerian derivatives of $\mathscr{L}_G$ and $\mathscr{L}_M$ with respect to h. We find:

$$\delta\mathscr{L}_G/\delta h_k{}^\mu = \Delta h^{k\nu}(R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R), \tag{61}$$

where $R_{\mu\nu}$ is the Ricci tensor. For $\mathscr{L}_M$ let us define $\theta_{\mu\nu}$ as

$$\delta\mathscr{L}_M/\delta h_k{}^\mu = -\Delta h^{k\nu}\theta_{\mu\nu}. \tag{62}$$

Then the field equations are

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = \theta_{\mu\nu}. \tag{63}$$

The left side is (i) symmetric in μ and ν, (ii) covariantly conserved. For consistency, $\theta_{\mu\nu}$ must also have these properties. As for symmetry; the local $SL(2, C)$ invariance of $\mathscr{L}_M$ leads to the identities

$$\theta_{\mu\nu} - \theta_{\nu\mu} = \frac{1}{2\Delta} b^j{}_\mu b^k{}_\nu \frac{\delta\mathscr{L}_M}{\delta\phi} S_{jk}\phi. \tag{64}$$

Therefore θ is symmetric when the matter field equations hold. As for covariant conservation: the fact that $\mathscr{L}_M$ is a scalar density under GCT, combined with occasional use of (64), gives further identities

$$\nabla_\mu \theta_\nu{}^\mu = \;\ldots\; \delta\mathscr{L}_M/\delta\phi \;\ldots \tag{65}$$

Thus the ϕ field equations are needed to make the gravitational ones algebraically and differentially consistent.

If the matter fields ϕ involve tensor representations of $SL(2, C)$ alone, it is easy to see that they can be replaced by new matter fields $\tilde{\phi}$ which are $SL(2, C)$ scalars but tensors under GCT matching the original $SL(2, C)$ properties of ϕ. The passage from ϕ to $\tilde{\phi}$ uses h and b, and is an admissible point transformation in the Lagrangian framework. When expressed in terms of $\tilde{\phi}$ rather than ϕ, it is clear that $\mathscr{L}_M$ does not involve h and A explicitly any more, but can be written exclusively in terms of $g_{\mu\nu}$, $\Gamma^{(0)\rho}_{\mu\nu}$ and $\tilde{\phi}$. Thus we recover in this case the metric theory. The Vierbein h and the $SL(2, C)$ gauge potential A are essential only if ϕ contains some spinorial components.

Option II This is based on viewpoint (a) of the previous section, and is the method proposed by Kibble. Even if there are no spinor matter fields, one finds slight differences from the standard metric theory.

One starts with 40 independent gravitational fields A^{ij}_{μ}, $h_k{}^{\mu}$. The gravitational Lagrangian density $\mathscr{L}_G$ is to be formed from the three tensors $R^{\beta}_{\alpha\mu\nu}$, $T^{\lambda}_{\mu\nu}$, $h_k{}^{\mu}$. Kibble chooses the expression of lowest degree:

$$\mathscr{L}_G = \tfrac{1}{2}\Delta R, \quad \Delta = (\det(h_k{}^{\mu}))^{-1},$$
$$R = h_i{}^{\mu} h_j{}^{\nu} F^{ij}_{\mu\nu} = g^{\alpha\nu} R^{\beta}_{\alpha\beta\nu}. \tag{66}$$

The total Lagrangian density, indicating the fields and derivatives that appear, is

$$\mathscr{L}_T = \mathscr{L}_M(\phi, \partial\phi, h, A) + \mathscr{L}_G(h, A, \partial A). \tag{67}$$

Note that the space-time derivatives of h are totally absent.

Let us introduce the notations

$$t^k{}_{\mu} = \frac{\partial \mathscr{L}_G}{\partial h_k{}^{\mu}} = \Delta \cdot b^j{}_{\mu} \cdot (R^{kl}{}_{jl} - \tfrac{1}{2}\delta^k_j R);$$
$$s^{\mu}_{ij} = \frac{\delta \mathscr{L}_G}{\delta A^{ij}_{\mu}} = -D_{\nu}[\Delta \cdot (h_i{}^{\mu} h_j{}^{\nu} - h_i{}^{\nu} h_j{}^{\mu})];$$
$$T^k{}_{\mu} = \frac{\partial \mathscr{L}_M}{\partial h_k{}^{\mu}}; \; S^{\mu}_{ij} = \frac{\partial \mathscr{L}_M}{\partial A^{ij}_{\mu}}. \tag{68}$$

Then the gravitational field equations are

$$t^k{}_{\mu} = -T^k{}_{\mu}, \; s^{\mu}_{ij} = -S^{\mu}_{ij}. \tag{69}$$

Now the invariance of $\mathscr{L}_G$ under local $SL(2, C)$ rotations leads to the identities

$$D_{\mu} s^{\mu}_{ij} = t_{ji} - t_{ij}, \; t_{ji} = h_i{}^{\mu} t_{j\mu}. \tag{70}$$

among the left sides of the field equations. Similarly, $\mathscr{L}_G$ being a scalar density under GCT leads to further identities

$$D_{\mu} t^{\mu}{}_{\nu} = \tfrac{1}{2} s^{\mu}_{ij} F^{ij}_{\nu\mu} - t^k{}_{\mu} D_{\nu} h_k{}^{\mu}. \tag{71}$$

The question again arises: do the expressions $T^k{}_{\mu}$, S^{μ}_{ij} derived from $\mathscr{L}_M$ have corresponding properties? Indeed they do. The local $SL(2, C)$ invariance and GCT scalar

density properties of $\mathscr{L}_M$ give us the identities

$$D_\mu S^\mu_{ij} = T_{ji} - T_{ij} - \frac{\delta \mathscr{L}_M}{\delta\phi} S_{ij}\phi,$$

$$D_\mu T^\mu{}_\nu = \tfrac{1}{2} S^\mu_{ij} F^{ij}_{\nu\mu} - T^k{}_\mu D_\nu h_k{}^\mu - \frac{\delta \mathscr{L}_M}{\delta\phi} D_\nu \phi. \tag{72}$$

So one is assured again that with the help of the matter field equations the gravitational ones become consistent.

The Euler–Lagrange field equations resulting from variations of the A field have a particularly interesting structure. They can be algebraically rewritten to express the torsion in terms of the object S^μ_{ij} arising from matter:

$$T^\lambda_{\mu\nu} = \Delta^{-1}(-S^\lambda_{\mu\nu} + \tfrac{1}{2}\delta^\lambda_\nu S^\sigma_{\mu\sigma} - \tfrac{1}{2}\delta^\lambda_\mu S^\sigma_{\nu\sigma}). \tag{73}$$

If there had been no matter fields at all, we see that the present theory of gravitation based on the fields A, h reduces to the metric theory! For in that case, $S^\lambda_{\mu\nu} = 0$ implies $T^\lambda_{\mu\nu} = 0$; then the affine connection Γ reduces to the Christoffel connection $\Gamma^{(0)}$; A^{ij}_μ takes up the form assumed for it in Option I; $R^\beta_{\alpha\mu\nu}$ reduces to the Riemann–Christoffel tensor; and the h-field equation becomes the vanishing of the Ricci tensor formed from the metric.

If there are matter fields, but if we assume that $L_M(\phi; \partial\phi)$ was at most linear in $\partial\phi$, simplifications still occur. For now it follows that S^μ_{ij} has no dependence on A^{ij}_μ. So the A-field equation (73) is a solution for torsion in terms of ϕ and h. We can use this in (56b) and so obtain an expression for A^{ij}_μ in the form

$$A^{ij}_\mu = f^{ij}_\mu(h; \partial h) + g^{ij}_\mu(h; \phi). \tag{74}$$

The f-term here is the same as in (58), and is just what A would have been if torsion had been zero.

Now: precisely because the field equations that have been used to solve for A^{ij}_μ are the equations

$$\delta\mathscr{L}_T / \delta A^{ij}_\mu = 0,$$

it is legitimate to put this solution (74) into $\mathscr{L}_T$ and regard the resulting expression as the Lagrangian for the fields ϕ and h. When this is done, we can separate the terms arising from f^{ij}_μ from those involving both f^{ij}_μ and g^{ij}_μ or g^{ij}_μ alone. The terms not involving g^{ij}_μ reproduce the total Lagrangian of Option I—as they must since f^{ij}_μ is what A^{ij}_μ would be if torsion were zero! Schematically,

$$\begin{aligned}
\mathscr{L}^{(\mathrm{II})}_T(\phi; \partial\phi; h; A; \partial A)|_{A=f+g} &= \mathscr{L}_M(\phi; \partial\phi; h; A)|_{A=f+g} \\
&\quad + \mathscr{L}^{(\mathrm{II})}_G(h; A; \partial A)|_{A=f+g} \\
&= \mathscr{L}^{(\mathrm{I})}_M(\phi; \partial\phi; h; \partial h) + \mathscr{L}^{(\mathrm{I})}_G(\text{Hilbert}) + \mathscr{L}'; \\
\mathscr{L}^{(\mathrm{I})}_M &= \mathscr{L}_M(\phi; \partial\phi; h; A)|_{A=f}, \\
\mathscr{L}^{(\mathrm{I})}_G(\text{Hilbert}) &= \mathscr{L}_G{}^{(\mathrm{II})}(h; A; \partial A)|_{A=f}; \\
\mathscr{L}' &= \text{remaining terms} \\
&= \text{direct } S\text{–}S \text{ (spin–spin) interactions.}
\end{aligned} \tag{75}$$

This effective Lagrangian for the independent fields ϕ, h can be directly compared to

the Lagrangian for these same fields under Option I. The only new term is $\mathscr{L}'$. But it occurs with one higher power of the gravitational coupling constant as compared to the remainder, and so from a practical point of view Option II reduces to Option I.

6. Gauging G and $\mathscr{P}$ simultaneously

In concluding this brief introduction to the gauge principle applied to the Poincaré group, we indicate how to combine the methods of §2 with those of §§3 and 4. Let the matter field Lagrangian density $L_M(\phi; \partial\phi)$ possess both G and $\mathscr{P}$ invariances. To gauge both groups and arrive at groups G', $\mathscr{P}'$, we must define a generalized derivative bringing in potentials A_μ^a, A_μ^{ij}, an affine connection $\Gamma_{\mu\alpha}^\beta$, and Vierbein fields h_k^μ. The ∇ operator is

$$\nabla_\mu = \partial_\mu + A_\mu^a T_a + \tfrac{1}{2} A_\mu^{ij} S_{ij} + \Gamma_{\mu\alpha}^\beta X_\beta^\alpha. \tag{76}$$

It preserves the G and $SL(2, C)$ representation properties of any field it acts on, and increases the GCT tensor behaviour by an extra covariant vector index. The commutator reads:

$$[\nabla_\mu, \nabla_\nu] = F_{\nu\mu}^a T_a + \tfrac{1}{2} F_{\nu\mu}^{ij} S_{ij} + R_{\alpha\mu\nu}^\beta X_\beta^\alpha + T_{\mu\nu}^\lambda \nabla_\lambda. \tag{77}$$

Here $F_{\mu\nu}^a$ is formed *exactly as in* §2 from A_μ^a; and $F_{\mu\nu}^{ij}$, $R_{\alpha\mu\nu}^\beta$, $T_{\mu\nu}^\lambda$ are as in §4. Imposing $\nabla h = 0$ leads to the same analysis as in §4 since h is G-invariant. The simplest Lagrangian density for the Yang–Mills field is evidently

$$\tfrac{1}{4}\Delta(h) g^{\mu\rho} g^{\nu\sigma} F_{\mu\nu}^a F_{a\rho\sigma}. \tag{78}$$

The transformation laws of A_μ^{ij} and $h_k{}^\mu$ are as before; for ϕ and A_μ^a we have

$$\delta\phi = (\varepsilon^a T_a + \tfrac{1}{2}\varepsilon^{ij} S_{ij})\phi - \xi^\nu \partial_\nu \phi,$$

$$\delta A_\mu^a = \varepsilon^b f_{bc}^a A_\mu^c - \partial_\mu \varepsilon^a - \partial_\mu \xi^\nu \cdot A_\nu^a - \xi^\nu \partial_\nu A_\mu^a. \tag{79}$$

The main point to appreciate is that there are three categories of fields: matter fields ϕ, Yang–Mills fields for internal symmetry, and gravitational fields. Their characteristic properties are shown by the following table:

	G'	$SL(2, C)'$	GCT
ϕ	Rep'n. with generators T_a	Rep'n. with generators S_{ij}	Scalars
A_μ^a	Gauge potential	Scalar	Cov't. vector
A_μ^{ij}	Scalar	Gauge potential	Cov't. vector
h_k^μ	Scalar	Vector	Contravariant vector

References

Utiyama R 1956 *Phys. Rev.* **101** 1597
Kibble T W B 1961 *J. Math. Phys.* **2** 212

Discussion

M. A. Melvin: You derived the symmetry of the energy momentum tensor from the requirement that the matter fields obey Lagrangian equations of motion. The usual derivation is given the requirement of conservation of angular momentum. Is there a correspondence between the two derivations?

N. Mukunda: In Lorentz invariant field theory the canonical energy momentum tensor may not be symmetric, but if we insist that it is such that the angular momentum density be the moment of the energy momentum density, then the latter is symmetric. With general covariance in a metric theory, the symmetry becomes automatic. With a violation it is again initially lost but later reestablished.

S. B. Khadkikar: Is there no coupling possible between Yang-Mill's fields and gravitational field? Especially when the former acquire mass though say spontaneous symmetry breaking is possible?

N. Mukunda: The coupling is there since the metric tensor has to be used in making the free Yang-Mill's Lagrange density a scalar density.

V. M. Rawal: You have discussed two options (a) and (b) and said if torsion is zero, we get the option (a). What is the difference then between the two at physical level?

N. Mukunda: It is in the spin-spin interaction which is very small. It is very difficult to measure experimentally.

Gauging the Lorentz group and long range interaction between spins

T PRADHAN
Institute of Physics, Sachivalaya Marg, Bhubaneswar 751 005, India

1. Formulation of the spin-gauge theory

The Sciama–Kibble (1961) approach of gauging the Poincare group which leads to the Einstein–Cartan theory of gravitation has the paradoxical feature of a contact spin–spin interaction emerging from a gauge theory. One of the motivations of gauge theories is to overcome the drawbacks of contact and other such non-normalizable interactions through the introduction of gauge fields. But here is an example of a gauge theory which gives rise to precisely such an interaction. Therefore a fresh look into this problem is necessary. For this it may be worthwhile to consider a gauge theory of the Lorentz group which is not clouded by the complications of gravity but at the same time gives rise to the spin–spin interaction. For this reason we choose a very specific co-ordinate dependence of the Lorentz parameters instead of the arbitrary co-ordinate dependence of these parameters as is done in the Sciama–Kibble approach of gauging of the Poincaré group.

We consider the Dirac Lagrangian

$$\mathscr{L}_0 = -\bar{\psi}(\gamma^\mu \partial_\mu + m)\psi, \tag{1}$$

for neutral spin 1/2 fermions. This Lagrangian is not invariant under the local Lorentz group. For the specific case where $\alpha_{\mu\nu}$ is such that (Naik and Pradhan 1981)

$$\tfrac{1}{6}\varepsilon^{\mu\nu\lambda\xi}\partial_\mu \alpha_{\nu\lambda}(x) = \partial^\xi \Lambda(x) \tag{2}$$

it is found that whereas $\mathscr{L}_0$ as given in (1) is not invariant under

$$\psi(x) \to \psi'(x') = \exp[\tfrac{i}{2}\alpha_{\mu\nu}(x)\Sigma^{\mu\nu}]\psi(x), \tag{3}$$

the modified Lagrangian

$$\mathscr{L}_D = -\bar{\psi}(\gamma^\mu D_\mu + m)\psi, \tag{4}$$

with

$$\begin{aligned}\gamma^\mu D_\mu \psi_\alpha &= \gamma^\mu(\delta_{\alpha\beta}\partial_\mu + \tfrac{ig}{4}\varepsilon^{\mu\nu\lambda\xi}\Sigma^{\alpha\beta}_{\nu\lambda}a_\xi)\psi_\beta \\ &= \gamma^\mu(\partial_\mu + \tfrac{3g}{4}\gamma_5 a_\mu)\psi_\alpha,\end{aligned} \tag{5}$$

where $a_\mu(x)$ is a massless axial vector boson, is invariant under the same transformation provided $a_\mu(x)$ transforms as

$$a_\mu(x) \to a'_\mu(x') = a_\mu(x) + \tfrac{2}{3}\alpha_{\mu\nu}(x)a^\nu(x) - \tfrac{1}{g}\partial_\mu \Lambda(x), \tag{6}$$

and its kinetic Lagrangian is of the form

$$\mathscr{L}_G = -\tfrac{1}{2}\partial_\mu a_\nu \partial^\mu a^\nu - \tfrac{3}{16}g^2(a_\mu a^\mu)^2. \tag{7}$$

with $a_\mu(x)$ satisfying the Lorentz condition

$$\partial^\mu a_\mu(x) = 0. \tag{8}$$

Combining (4) and (7) we get for the total Lagrangian of the neutral Dirac field and gauge field

$$\begin{aligned}\mathscr{L} = & -\bar{\psi}(\gamma^\mu \partial_\mu + m)\psi - \tfrac{1}{2}\partial_\mu a_\nu \partial^\mu a^\nu \\ & - \tfrac{3}{4} g \bar{\psi}\gamma^\mu \gamma_5 \psi a_\mu - \tfrac{3}{16} g^2 (a^\mu a_\mu)^2.\end{aligned} \tag{9}$$

We shall name the axial vector gauge particle as axial photon and the resulting gauge theory as spin gauge theory. Since photon has spin it will also couple to the gauge field $a_\mu(x)$. The interaction Lagrangian can be obtained by using covariant derivative

$$D_\mu A_\gamma = \partial_\mu A_\gamma + \tfrac{ig}{4}\varepsilon_{\mu\nu\lambda\zeta}\Sigma^{rs}_{\nu\lambda} a_\zeta A_s \tag{10}$$

in the photon Lagrangian

$$\mathscr{L}_\gamma = -\tfrac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) \tag{11}$$

and is found to be

$$\mathscr{L}_{\text{int}} = -g\varepsilon^{\mu\nu\lambda\zeta}\partial_\mu A_\nu A_\lambda a_\zeta - \tfrac{1}{2}g^2(a^\mu a_\mu A^\nu A_\nu - A^\mu a_\mu A^\nu a_\nu). \tag{12}$$

2. Long-range spin–spin force

The two-body spin–spin force resulting from the spin-gauge theory formulated above can be obtained to lowest order perturbation from Feynman diagrams of figure 1. Suitable non-relativistic limit leads to potential

$$v_{ab}(\mathbf{r}) = -\frac{g^2}{2r}\left[\boldsymbol{\sigma}_a \cdot \boldsymbol{\sigma}_b + \frac{(\boldsymbol{\sigma}_a \cdot \mathbf{r})(\boldsymbol{\sigma}_b \cdot \mathbf{r})}{r^2}\right] \tag{13}$$

between two spin 1/2 particles. Similar calculation leads to the potential

$$v_{ab}(\mathbf{r}) = -\frac{g^2}{2r}\left[\mathbf{S}_a \cdot \mathbf{S}_b + \frac{(\mathbf{S}_a \cdot \mathbf{r})(\mathbf{S}_b \cdot \mathbf{r})}{r^2}\right] \tag{14}$$

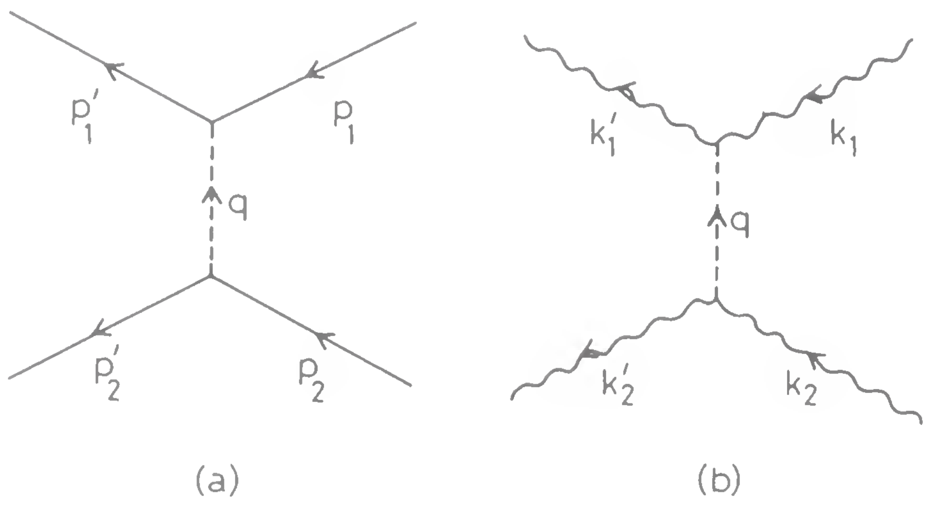

Figure 1. Feynman diagrams

between two photons and the potential

$$v_{\nu\gamma}(\mathbf{r}) = -\frac{3g^2}{8r}\left[\sigma\cdot\mathbf{S} + \frac{(\sigma\cdot\mathbf{r})(\mathbf{S}\cdot\mathbf{r})}{r^2}\right] \tag{15}$$

between photon and four-component neutrino. It is clear from (13), (14) and (15) that the force between two particles with spins is of long range and Coulomb-like. It is attractive for parallel spins and repulsive for antiparallel ones. This is precisely the experimental result of Tam and Happer (1977) who demonstrated that while circularly-polarised laser beams of opposite polarization repel each other, beams with same polarization attract in a medium of sodium vapour for laser frequencies on the high frequency wing of the D_1 line. It is therefore natural to interpret this repulsion attraction between laser beams in terms of the long range spin–spin force. This long-range spin-spin force must be very weak. Otherwise it would contribute substantially to electromagnetic processes such as Moller scattering, Compton scattering etc. Since our claim is that this super weak interaction is responsible for the observed attraction and repulsion between polarized laser beams in sodium vapour, it is necessary to explain the mechanism of its enhancement due to spin polarized sodium atoms. The circularly polarized laser beams with a frequency slightly higher than that of the sodium D_1 line optically pumps the atoms from the ground state to $3^2p_{1/2}$ state, thereby producing a medium of spin-polarized atoms. The interaction of the spins of the laser beams gets enhanced by this medium. A many-body formulation of this enhancement and comparison with experimental results of Tam and Happer (1977) leads to $g^2/4\pi = 0{\cdot}7 \times 10^{-9}$.

3. Comparison with Sciama–Kibble approach

One important point of difference between Sciama–Kibble Poincare gauge theory and any local internal symmetry group theory is that while the invariance of the Lagrangian density ensures the invariance of the action function in the latter, this is not the case in the former. This can be seen as follows:

$$S = \int d^4x\,\mathscr{L}(x) \rightarrow S' = \int d^4x'\,\mathscr{L}'(x') = \int d^4x[1 + \partial_\mu(\delta x^\mu)]\,\mathscr{L}'(x')$$
$$= \int d^4x\,\mathscr{L}'(x). \tag{16}$$

Thus invariance of action function *i.e.*

$$S' = S$$

requires

$$\mathscr{L}(x) = \mathscr{L}'(x), \tag{17}$$

but not

$$\mathscr{L}(x) = \mathscr{L}'(x'), \tag{18}$$

as in the case of local internal symmetry gauge theories. Equation (17) implies that in the Sciama–Kibble approach one considers a group of transformations in the local tangent

space where the spinor fields transform as

$$\psi'(x) = \exp[\tfrac{i}{2}\alpha_{\mu\nu}(x)\Sigma^{\mu\nu}]\,\psi(x). \tag{19}$$

For this reason vierbeins are required in the definition of covariant derivatives. On the other hand in our approach since we have chosen

$$\begin{aligned} &\partial_\mu \alpha_{\nu\lambda} = \varepsilon_{\mu\nu\lambda\zeta}\partial_\zeta \Lambda(x), \\ &\partial_\mu(\delta x^\mu) = \partial_\mu(\alpha^{\mu\nu}x_\nu) = (\partial_\mu \alpha^{\mu\nu})x_\nu = 0, \end{aligned} \tag{20}$$

it follows from (16) that

$$\mathscr{L}'(x') = \mathscr{L}(x) \tag{21}$$

for action function to be invariant under the group transformations. Thus while in the Sciama–Kibble theory $\mathscr{L}(x) = \mathscr{L}'(x)$, in our spin gauge theory, $\mathscr{L}(x) = \mathscr{L}'(x')$, and no vierbeins are necessary.

4. Consequences of long range spin–spin interaction

Several interesting consequences follow from the long-range attractive force between parallel spins. One obvious result is the formation of hydrogen-like bound states between neutron and electron. Since the force is very weak (weaker by a factor of 10^{-7} compared to electromagnetic force) such bound states would have a large size of the order of few millimeters and electromagnetic transitions between its discrete states would give rise to radiations of few meters wavelength. Bound states between electron and neutrino as well as between photons would also be possible. Further, just as electron acquires a charge distribution on account of its interaction with the photon, a spinning particle would acquire a spin distribution on account of its interaction with the axial photon. Because of the weakness of the coupling constant, this distribution would be limited to a radius of the order of 10^{-19} cm.

On account of the fact that parallel spins attract in contrast to like charges repelling, a spinning particle placed in a spin polarizable medium would be antiscreened rather than being screened like a charged particle in a charge-polarizable medium. As a result, the effective coupling constant for spin–spin interaction would increase with distance. In other words the spin gauge theory is an asymptotically free field theory.

The attraction between parallel spin has one very attractive feature in that it can provide stability to the classical electron whose charge distribution would ordinarily fly apart on account of repulsion between like charges. The spin distribution on the other hand would provide a counter force to hold together the parts of the electron.

The coupling of the spin to the axial photon would not affect observable processes of quantum electrodynamics on account of the exceedingly small value of the coupling constant. However, this coupling when extrapolated to the electron's charge radius could become as large as the electromagnetic coupling constant. In such an event the divergent part of electron self-energy resulting from emission followed by reabsorption of photon could cancel with that arising from emission and reabsorption of axial photon. Preliminary calculation using Feynman diagrams in the quantized spin-gauge theory where ghosts make their appearance, shows that all divergences do cancel if the coupling constant is suitably extrapolated so that it equals the electromagnetic coupling constant.

5. Summary and conclusions

We thus conclude that there exists an elementary universal interaction between spins of particles which follow from a gauge principle. This force is of long range and weaker than the weak interaction. It is of the same order of magnitude as the super-weak interaction. This force can provide stability to the otherwise unstable classical electron and can remove all divergences in quantum electrodynamics without disturbing its present excellent agreement with experiment. It can give rise to loose bound states of very small binding energy between neutral particles possesing spin thereby opening up a new frontier of very long wavelength spectroscopy.

References

Kibble T W B 1961 *J. Math. Phys.* **2** 212
Naik P C and Pradhan T 1981 *J. Phys.* **A14** 2795
Sciama D W 1962 in *Recent developments in general relativity* (Oxford Pergamon)
Tam A C and Happer W 1977 *Phys. Rev. Lett.* **38** 278

Discussion

T. Padmanabhan: (a) Is the theory with long range spin-spin interaction renormalizable inspite of only partial gauging of Lorentz group being resorted to? (b) What is the physical meaning of the set of transformations? (c) Is there a group out of these transformations? What is its structure?

T. Pradhan: (a) Yes, the theory is renormalizable. It is a full-fledged gauge theory. There is no such thing as partial gauging. (b) The physical meaning is that spin is a dynamical entity in this gauge theory which can be measured with the aid of the gauge field instead of being an algebraic entity as it is in the global group. (c) Yes, these transformations form a group, a local gauge group.

N. Mukunda: What subgroup of SL (2, c) has been gauged, or is it the whole group?

T. Pradhan: As far as I can see, the whole group has been gauged.

P. Majumdar: What about Adler–Bell–Jackiw anomalies for the axial photon a_μ?

T. Pradhan: As a matter of fact the Adler–Bell–Jackiw anomaly can be cured through the introduction of the photon-axial photon interaction as the divergence of the axial vector current in question has the same structure as in the Adler–Bell–Jackiw case.

M. A. Melvin: Could circular birefringence (different behaviour of right and left circularly polarised in the medium) induced by the laser beam be the explanation of separation of the two components in the experiment? If so then the reduction of the effect by a transverse magnetic field would be interpreted as a reduction of the birefringence

T. Pradhan: I am unable to see how it can be due to birefringence.

K. Subramanian: If you have $g = e$ you have said that vacuum self-energy graphs involving axial photons and real photons cancel and make the theory finite. Would this

not lead to wrong predictions of the Lamb shift (which is very accurately measured)? Therefore one does not really have a finite theory since $g = e$ is not borne out experimentally.

T. Pradhan: Since the theory is asymptotically free one could renormalize at a momentum transfer where $g = e$. The value of g that would occur in Lamb-shift calculations need not be at the same momentum transfer. The details of this scheme of making self-energies finite have not yet been worked out.

N. Dadhich: Should one take that $e = g$ is just a formal and hypothetical relation so as to cancel out divergences in self energy?

T. Pradhan: Comparison with laser experiment shows that $g^2/4\pi \cong 10^{-9}$. However since the theory is asymptotically free it would be possible to get to a subtraction point where $g = e$. However, the details of this have not yet been worked out.

A. K. Raychaudhuri: It is imagined that the electron is stable under electrical interaction and spin-spin interaction. Should we expect that the theory will give the numerical value of the fine structure constant from the condition of equilibrium of the electron.

T. Pradhan: Since a new coupling constant g has appeared in the theory the stability considerations can only fix this coupling constant with respect to the fine-structure constant. It cannot give a numerical value of the fine structure constant.

Introduction to supergravity

PARTHASARATHI MAJUMDAR
Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India
Present address: Saha Institute of Nuclear Physics, 92, APC Road, Calcutta 700 009, India

1. Introduction

The strong, weak and electromagnetic interactions describing elementary particle processes at energies far below the Planck mass, have been formulated as a (grand) unified theory (Langacker 1981) within the framework of local internal non-Abelian gauge theories. Gravitation, which corresponds to a space-time (rather than an internal) symmetry, has also received an adequate treatment within the non-Abelian gauge theory idea (Kibble 1960). However, the incorporation of both space-time and internal symmetries within a single unifying framework has been a formidable problem for a long time. Indeed, the stumbling block for such unification in flat space is a rigorous 'no-go' theorem, due to Coleman and Mandula (1967): Given a Lie group G which is the maximal symmetry group of the S-matrix, such that (i) it has a subgroup G' locally isomorphic to P, the Poincaré group, (ii) it has finite dimensional irreducible representations with mass M (a bounded spectrum), (iii) the S-matrix is an analytic function of s (c.m. momentum squared) and t (momentum transfer squared), (iv) the S-matrix is non-trivial except at some values of s; then, G is locally isomorphic to $P \times I$, where I is some internal symmetry Lie group with all its generators commuting with those of P. Thus, this theorem forbids any mixing of internal and space-time symmetries.

The proof of the Coleman-Mandula theorem depends crucially on the Lie property of the infinitesimal algebra of G. Thus as long as G is a Lie group, gravitation cannot be unified with the other interactions within the philosophy of gauge theories. One has, at this point, the option of either attempting unification from a very different standpoint, or considering more general algebraic structures for the maximal symmetry group G in order to sidestep the Coleman-Mandula no-go theorem. The latter option is exercised in theories with *supersymmetry* (Fayet and Ferrara 1977; Wess and Bagger 1981), where the infinitesimal generator algebra of G is a *graded* Lie algebra, whose 'even' elements obey the usual commutation rules of a Lie algebra, but which has, in addition, a set of 'odd' elements which have anti-commutation relations between themselves and commutation relations with the 'even' generators. This larger algebra circumvents the restrictions of the CM theorem *vis-a-vis* mixing of internal and space-time symmetries, and opens up possible avenues of unification involving gravitation. Acting on local fields this algebra has the effect of transforming a fermion field into a boson field and vice-versa, hence the name supersymmetry.

In what follows, we shall first of all discuss briefly, the supersymmetry algebra obtained by grading the Poincaré Algebra, its extensions and irreducible representations, in flat space (global supersymmetry). We shall point out the action of this

algebra on local fields, and exhibit local field theories invariant under global supersymmetry transformations. One then generalizes these to the case where the supersymmetry (ss) transformations depend on space-time (local supersymmetry). It turns out that successive local ss transformations are equivalent to a *general coordinate transformation*. Thus, local ss is intimately linked to gravitation. Furthermore, the nature of the ss transformation implies a Noether current transforming as a spin-3/2 field under homogeneous Lorentz transformations, thus requiring a spin 3/2 gauge field (gravitino) in the locally supersymmetric (gauged) version. The locally supersymmetric spin 2-spin 3/2 theory, called simple supergravity, is written down as the sum of the Rarita-Schwinger and Einstein-Cartan actions. We next discuss the consistency of the equations of motion of this system in the light of the new fermionic gauge symmetry. This is followed by brief descriptions of the matter-couplings of the simple supergravity multiplet, as well as its extensions (N-extended supergravity). We then go on to a short exposé of the Hamiltonian structure of simple supergravity, alluding to the beautiful work of Teitelboim (1977) and Deser and Teitelboim (1977) and Grisaru (1978) which has led to (a) an alternative derivation of the supergravity action in terms of the Dirac square root of the Einstein theory, (b) a proof of the positivity of the ADM mass in classical Einstein theory. Next we introduce superspace and briefly discuss the various formulations of supergravity in superspace. We conclude with a few remarks on the renormalizability properties of quantized supergravity.

2. Global supersymmetry

We augment the Poincaré algebra

$$[P_\mu, P_\nu] = 0; \quad [P_\mu, M_{\rho\lambda}] = [P_\rho \eta_{\mu\lambda} - P_\lambda \eta_{\mu\rho}] \tag{1}$$

$$[M_{\mu\nu}, M_{\rho\lambda}] = M_{\mu\rho}\eta_{\nu\lambda} + \text{three other terms},$$

where P_μ, $M_{\rho\lambda}$ are the translation generators and generators of homogeneous Lorentz transformations respectively, by introduction of a new Majorana spinor generator Q_α, $\alpha = 1, 2, 3, 4$, obeying,

$$[Q_\alpha, P_\mu] = 0, \; [Q_\alpha, M_{\mu\nu}] = \tfrac{1}{2}(\sigma_{\mu\nu})_{\alpha\beta} Q_\beta, \tag{2}$$

where $\sigma_{\mu\nu} \equiv \frac{1}{4}[\gamma_\mu, \gamma_\nu]$. Further, we postulate that

$$\{Q_\alpha, \bar{Q}_\beta\} = (\gamma^\mu)_{\alpha\beta} P_\mu, \tag{3}$$

where, in the Majorana representation for γ-matrices, $\bar{Q}_\beta = (Q^T\gamma^0)_\beta$. Note that (3) contains an anti-commutator, as it must, in order to be consistent with (2); this is the novel aspect of the algebra described by (1) to (3).

Two consequences of this 'super' algebra are immediately evident:

(i) Since $[Q_\alpha, P^2] = 0$, members of the same irreducible representation of the algebra have the same mass. But, because $[Q_\alpha, M_{\mu\nu}] \neq 0$, particles of different spin belong to the same (super) multiplet. This latter aspect is specific to the supersymmetry algebra and has no counterpart in ordinary internal symmetry algebras.

(ii) Equation (3) implies that successive actions of Q_α are equivalent to a coordinate translation; or in more picturesque language, 'a supersymmetry transformation is the square root of a translation'.

One can generalize to the case of several (N) Q's, labelled Q^i_α, $i = 1, \ldots N$, satisfying the algebra

$$[Q^i_\alpha, P_\mu] = 0; \quad [Q^i_\alpha, M_{\mu\nu}] = \tfrac{1}{2}(\sigma_{\mu\nu})_{\alpha\beta} Q^i_\beta,$$
$$\{Q^i_\alpha, Q^i_\beta\} = \delta^{ij}(\gamma_\mu)_{\alpha\beta} P^\mu + (\mathbf{1})_{\alpha\beta} Z^{ij} + (\gamma_5)_{\alpha\beta} \tilde{Z}^{ij}, \tag{4}$$

where Z, $\tilde{Z}$ are called central charges; they commute with all generators of the superalgebra. The Q^i's have a natural N-dimensional (vector) representation under $O(N)$

$$[Q^i_\alpha, B_I] = S^{ij}_I Q^j_\alpha; \quad [B_I, B_J] = f_{IJK} B_K, \tag{5}$$

where B_I are the generators of the internal group $O(N)$. Equations (4) and (5) provide the first glimpses of non-trivial mixing between space-time and internal symmetries in flat space.

Consider now, the lowest dimensional irreducible representations of the super Poincaré algebra. For simplicity, we restrict ourselves to the massless case. The supermultiplets are characterised in this case by the helicity. In the simplest case with only one Q_α ($N = 1$ ss) the supermultiplet with lowest dimensionality in terms of helicity states (4) has the helicity content $(\pm\lambda, \pm(\lambda - \frac{1}{2}))$. For example, with $\lambda = 2$, we have the multiplet $(\pm 2, \pm 3/2)$, which is the celebrated simple supergravity supermultiplet. $\lambda = 1/2$ gives us the supermultiplet $(\pm 1/2, \pm 0)$, which is the Wess-Zumino (1974) multiplet. In the case of extended supersymmetry, the Q's and $\bar{Q}$'s act as lowering and raising operators, and different helicity states with various multiplicities depending on N appear in the irreducible supermultiplet of lowest dimension. We reproduce in table 1 various multiplicities for $O(N)$ ss due to Scherk (1979).

Table 1. Multiplicities of various supermultiplets in N-extended ss.

Spin \ N	1				2				3			4			5	6	7	8	9
5/2																			1
2	1			1				1			1			1	1	1	1	1	10
3/2			1	1			1	2		1	3		1	9	5	6	8	8	45
1		1	1			1	2	1	1	3	3	1	4	6	10	16	28	28	120
1/2	1	1			2	2	1		4	3	1	4	7	4	11	26	56	56	210
0	2				4	2			6	2		6	8	2	10	10	70	70	256

This shows that for $N > 8$, the lowest dimensional supermultiplet has 10 gravitons and a spin 5/2 particle. In view of the recently found inconsistencies (Aragone and Deser 1979) in the couplings of the spin 5/2 field and also from the aesthetic requirement of a *single* graviton, $N = 8$ ss appears to be the largest extended ss that need be considered.

We now discuss the action of the supersymmetry algebra, in particular of the supersymmetry generator Q, on local fields. These would yield field-theoretic realizations of the algebra (Fayet and Ferrara 1977, Wess and Bagger 1981).

2.1 *Wess-Zumino multiplet* [1/2, 0]:

This model contains a scalar field $A(x)$, a pseudoscalar $B(x)$ and a Majorana spinor $\psi(x)$. The kinetic energy part of the action is

$$I_0 = \int d^4x \left\{ \tfrac{1}{2}(\partial\mu A)^2 + \tfrac{1}{2}(\partial\mu B)^2 + \frac{i}{2}\bar{\psi}\not{\partial}\psi \right\}. \tag{6}$$

Under the supersymmetry transformations

$$\delta_s A \equiv [Q, A] = i\bar{\varepsilon}\psi; \quad \delta_s B = i\bar{\varepsilon}\gamma_5\psi,$$
$$\delta_s\psi \equiv \{Q, \psi\} = \partial_\mu(A + \gamma_5 B)\gamma^\mu\varepsilon, \tag{7}$$

where ε is a space-time independent totally anticommuting Majorana 4-spinor parameter, one obtains $\delta_s I_0 = 0$. The Noether (super)-current corresponding to this invariance is

$$S_{\mu\alpha} = [\partial_\nu(A + \gamma_5 B)\gamma^\nu\gamma_\mu\psi]_\alpha \tag{8}$$

2.2 *Vector multiplet*:

(1, 1/2) [contains a vector and Majorana spinor]

$$I_0 = \int \mathrm{d}^4x\{-\tfrac{1}{4}F^2_{\mu\nu} + \tfrac{i}{2}\bar{\lambda}\not{\partial}\lambda\}, \tag{9}$$

$\delta_s I_0 = 0$ under

$$\delta_s A_\mu = i\bar{\varepsilon}\gamma_\mu\lambda, \; \delta_s\lambda = \sigma_{\mu\nu}F^{\mu\nu}\varepsilon, \tag{10}$$

$$s_{\mu\alpha} = 2F^{\rho\nu}(\sigma_{\rho\nu}\gamma_\mu\lambda)_\alpha. \tag{11}$$

2.3 *Linearized supergravity*:

(2, 3/2) [symmetric 2nd rank tensor and vector spinor]

$$\begin{aligned} I = I_2 + I_{3/2} = \int \mathrm{d}^4x\{[(\eta^{\mu\rho}h^{\nu\lambda} - \eta^{\mu\lambda}h^{\nu\rho}) \\ \times(\partial_\mu\omega_{\nu\rho\lambda} - \partial_\nu\omega_{\mu\rho\lambda}) + (\omega^\sigma_{\mu\rho}\omega_{\nu\sigma\lambda} - (\rho\leftrightarrow\lambda)) \\ \times\eta^{\mu\nu}\eta^{\rho\sigma}] - \tfrac{i}{2}\varepsilon^{\mu\nu\rho\sigma}\bar{\psi}_\mu\gamma_5\gamma_\nu\partial_\rho\psi_\sigma\}, \end{aligned} \tag{12}$$

where the last term is the Rarita–Schwinger action. $\delta_s I = 0$ under

$$\delta_s h^{\mu\nu} = \tfrac{1}{2}\bar{\varepsilon}(\gamma^\mu\psi^\nu + \gamma^\nu\psi^\mu); \; \delta_s\psi_\mu = \tfrac{1}{2}\omega_{\mu\nu\lambda}(h)\sigma^{\nu\lambda}\varepsilon, \tag{13}$$

where $\omega_{\mu\nu\lambda}(h) = \Sigma(\partial h)$ [schematically],

$$s_{\mu\alpha} \cong \tfrac{i}{4}\varepsilon_{\mu\nu\rho\lambda}[\gamma_5\gamma^\nu\psi^\rho\omega^{\lambda\sigma\varepsilon}(h)\sigma_{\sigma\varepsilon}]_\alpha. \tag{14}$$

Additional invariances:

$$\delta\psi_\mu = \partial_\mu\alpha(x) \text{ [Abelian gauge invariance]}, \tag{15}$$

$$\delta h_{\mu\nu} = \partial_\mu\xi_\nu(x) \text{ [Linearized co-ordinate invariance]}.$$

One observes from these, *e.g.*, from (a) that

$$[\delta_s^1, \delta_s^2] A(x) = 2\bar{\varepsilon}_2 \gamma^\mu \varepsilon_1 \partial_\mu A(x). \tag{16}$$

This is referred to as 'the square root property' of ss.

3. Local supersymmetry

We now generalize to the case where $\varepsilon = \varepsilon(x)$, *i.e.* gauge the supersymmetry algebra locally. An immediate consequence emerges from (16)

$$[\delta_s^1, \delta_s^2] A(x) = 2\bar{\varepsilon}_2(x) \gamma^\mu \varepsilon_1(x) \partial_\mu A(x) + \text{other terms}. \tag{17}$$

The quantity $2\bar{\varepsilon}_2(x)\gamma^\mu\varepsilon_1(x) \equiv a^\mu(x)$, indicates a space-time dependent translation, also known as a general coordinate transformation. (The 'other terms' exist for local $\varepsilon(x)$ but need not concern us here). It follows that local ss is intimately linked to gravity, or, in particular, the Einstein action of general relativity must be part of *every* locally supersymmetric action. This implies further that the simplest (minimal) locally supersymmetric model must also contain the supersymmetric partner of the graviton, namely the spin 3/2 gravitino. This model is called simple (or $N = 1$) supergravity.

There are several alternative derivations of the simple supergravity action. We outline here the approach of Deser and Zumino (1976). They use the tetrad formalism where one introduces tetrad (or vierbein) fields $e_\mu^a(x)$, $e_a^\mu(x)$, (where Greek indices refer to the general Riemannian manifold, and Latin ones, to the tangent Minkowski space at point x) satisfying the relations

$$e_\mu^a e_{\nu a} = g_{\mu\nu}, \quad e_\mu^a e^{\mu b} = \eta^{ab}.$$

General coordinate frame (world) tensors are linked to tangent space tensors by means of these tetrads. Further, they use the first order formalism for gravitation (Kibble 1960), involving the Einstein–Cartan action for gravitation where $e_\mu^a(x)$ and the spin connection functions $\omega_\mu^{ab}(x)$ are treated as independent variables. The gravitational part of the action is therefore,

$$I_g[e, \omega] = \int d^4x\, e(x) e_a^\mu(x) e_b^\nu(x) R_{\mu\nu}^{ab}(x), \tag{18}$$

where $e \equiv \det e_\mu^a$ and $R_{\mu\nu}^{ab}(\omega) \equiv \partial_\mu \omega_\nu^{ab} + \omega_{\mu c}^a \omega_\nu^{cb} - (\mu \leftrightarrow \nu)$. The gravitino is described in flat space by the Rarita–Schwinger action (1941); we use here the minimal coupling prescription to couple it to gravitation:

$$\partial_\mu \psi_\rho \to D_\mu \psi_\rho \equiv \partial_\mu \psi_\rho + \tfrac{1}{2} \omega_\mu{}^{ab} \sigma_{ab} \psi_\rho. \tag{19}$$

Thus,

$$I_{3/2} = -\tfrac{1}{2} \int d^4x\, \varepsilon^{\mu\nu\rho\lambda} \bar{\psi}_\mu \gamma_5 \gamma^a e_{\nu a} D_\rho \psi_\lambda, \tag{20}$$

$$I_{SG} = \int d^4x \{ e e_a^\mu e_b^\nu R_{\mu\nu}^{ab} - \tfrac{1}{2} \varepsilon^{\mu\nu\rho\lambda} \bar{\psi}_\mu \gamma_5 \gamma^a e_{\nu a} D_\rho \psi_\lambda \}. \tag{21}$$

Equation (21) is the celebrated supergravity action in the form first proposed by Deser and Zumino (1976)

An alternative derivation has been given by Boulware *et al* (1979), extending the analysis of Deser (1970) for pure gravity. In this scheme, the invariances of linearized supergravity, *viz* global ss, local Abelian gauge invariance and linearized general coordinate invariance, as shown to imply, by self-consistency arguments, the full nonlinear action of supergravity. In a third approach based on the second order formalism $[\omega = \omega(e) + \kappa(\psi)]$, Freedman *et al* (1976) have also derived the complete supergravity action (for reviews of these methods, see van Nieuwenhuizen 1981 and Deser 1980). This last approach was historically perhaps the earliest derivation of the supergravity action.

The field equations (in first order form) for (21) are

$$\delta e\colon G^{\mu a} \equiv R^{\mu a} - \tfrac{1}{2} e^{\mu a} R = \underbrace{\tfrac{i}{2} e^{-1} \varepsilon^{\mu\nu\alpha\beta} \overline{\psi}_\nu \gamma_5 \gamma^a D_\alpha \psi_\beta}_{T^{\mu a}(\psi)}, \tag{22}$$

$$\delta\omega\colon \underset{(\text{torsion})}{C^a_{\nu\rho}} \equiv D_{[\nu} e^a_{\rho]} = \tfrac{i}{2} \overline{\psi}_\nu \gamma^a \psi_\rho, \tag{23}$$

$$\delta\psi\colon R^\mu \equiv \varepsilon^{\mu\nu\alpha\beta} \gamma_5 \gamma_a (e^a_\nu D_\alpha \psi_\beta - C^a_{\nu\alpha} \psi_\beta) = 0. \tag{24}$$

Given the form of (24), one raises the question as to whether $R^\mu = 0$ is valid when (22) and (23) are used. This is not an idle question since precisely this issue of consistency has thwarted all previous attempts to couple the Rarita–Schwinger field to other fields like the Maxwell field.

To study this question, one calculates $D_\mu R^\mu \equiv D_\mu (\delta I_{SG}/\delta\psi_\mu)$ and inserts (22) and (23) to determine if $D_\mu R^\mu \equiv 0$. [In the Rarita–Schwinger Maxwell system, *e.g.* $D_\mu R^\mu \neq 0$, see Deser 1980 and references therein]. In this case one can show that

$$\delta I_{SG} \equiv \left\{ \frac{\delta I}{\delta e^a_\mu} \delta e^a_\mu + \frac{\delta I}{\delta \psi_\mu} \delta\psi_\mu + \frac{\delta I}{\delta \omega^{ab}_\mu} \delta\omega^{ab}_\mu \right\} = 0. \tag{25}$$

under the local ss transformations

$$\begin{gathered} \delta_s \psi_\mu = 2 D_\mu \varepsilon(x); \quad \delta_s e^a_\mu(x) = i\bar{\varepsilon}(x) \gamma^a \psi_\mu(x) \\ \delta_s \omega^\mu_{ab} = \varepsilon^{\mu\nu\lambda\rho} \bar{\varepsilon}(x) e_{\nu b} \gamma_5 \gamma_a D_\lambda \psi_\rho. \end{gathered} \tag{26}$$

This implies that as a result of (22) and (23) and Fierz transformations of the Majorana spinor,

$$D_\mu R^\mu \equiv 0. \tag{27}$$

Thus, the consistency of the Rarita–Schwinger Einstein system is ensured by local ss. Additional features of the supergravity action are of course general coordinate invariance and local Poincaré invariance. The Abelian invariance of ψ_μ is now replaced by a non-Abelian fermionic gauge invariance, [*i.e.* $\alpha(x) = \varepsilon(x)$], with ψ_μ acting as the gauge field. It is remarkable that this new fermionic invariance has (a) made all ψ_μ couplings consistent, and (b) fixed the minimal coupling strength at κ. However, the supermultiplet (2, 3/2) is to be viewed as a gauge multiplet, and one must therefore consider coupling of (globally) supersymmetric lower spin matter to this multiplet.

4. Matter couplings and extended supergravity

In the above, we saw that the local supersymmetry algebra closes only when the field equations are used. For off-shell invariance, one must introduce non-dynamical auxilliary fields which 'uniformize' the ss transformations. For the simple supergravity theory, we need a scalar $S(x)$, a pseudoscalar $P(x)$ and a vector $V(x)$, to yield closure of the algebra without using field equations. The auxilliary Lagrange density is

$$\mathscr{L}_{\text{Aux}} = \tfrac{1}{2}(S^2 + P^2 + V_\mu V^\mu) + O\cdot\text{T}. \tag{28}$$

Thus, the absence of kinetic energy terms makes these fields unpropagating. These turn out to be useful for matter coupling. In general, the (2, 3/2) system is coupled to 'super' matter minimally through the supercurrent $s_{\mu\alpha}^{\text{matter}}$: $\mathscr{L}_{\text{matter}} = g\psi_{\mu\alpha}s_\alpha^\mu$. Local ss also requires contact terms analogous to $e^2|\varphi|^2 A_\mu^2$ in scalar QED. This implies that the ss transformation of ψ_μ is modified: $\delta\psi_\mu = D_\mu\varepsilon(x) +$ matter-dependent terms. Introduction of S, P and V_μ fields makes these additional terms tractable. However, so far there is no geometrical understanding of these auxiliary fields. Only their algebraic role turns out to be useful.

Let us now consider extensions of the simple supergravity theory, *i.e.* systems with more than one global charge. From table 1, it is obvious that the number of gravitinos in any $O(N)$ supergravity theory is equal to N, the number of spinorial generators Q^i. Thus, take for example

$$\begin{aligned} I &= I_{SG} + I(\tfrac{3}{2}, 1), \\ &= I_E + I_{3/2}(\psi) + I_{3/2}(\chi) + I_1(A) + \textstyle\int \mathrm{d}^4x\overline{\psi}\cdot\mathscr{S} + I_{\text{coup}}. \end{aligned} \tag{29}$$

One can show that there exists a global $O(2)$ symmetry: $\psi_\mu^1 \equiv \psi_\mu$, $\psi_\mu^2 \equiv \chi_\mu$ transform vectorially under $O(2)$, and leaves I invariant. One can define a *Dirac* gravitino $\psi_\mu^D \equiv \psi_\mu + i\chi_\mu$: this just shows that a Dirac spinor is an $O(2)$ object. An immediate outcome of this above example is the consistent coupling of the gravitino and Maxwell fields, which has been impossible to achieve outside of ss (Deser 1980).

By inclusion of more lower spin fields in the above manner, one can construct consistently-coupled locally supersymmetric theories of extended supergravity, corresponding to 3, 4, . . . generators. The maximal extended supergravity model consistent with a single graviton and no higher (*i.e.* > 2) spin fields is the $N = 8[O(8)]$ supergravity theory. Since $O(8) \not\supset SU(3)_C \times [SU(2) \times U(1)]_{W-S}$ it is not a phenomenologically useful theory for unification of all particle interactions, if treated as a fundamental theory (*i.e.* if $O(8)SG$ fields are considered as the fundamental fields). However, Ellis (1981) has recently proposed a scenario in which the $O(8)$ supergravity theory is a theory of preons of which the known particles (gluons, W-bosons, quarks etc.) are bound states. In this case it is possible in principle to achieve 'supergrand unification'. However, the theory is too unwieldly to base any meaningful predictions on, as yet.

5. Hamiltonian structure of supergravity

In this section, our aim is to delineate the additional features of the canonical constraint structure of simple supergravity, with regard to that of gravity. First of all, observe that

in flat space, the Rarita–Schwinger action can be written non-covariantly as

$$I_{3/2} = \tfrac{1}{2}\int d^4x\, \varepsilon^{oijk}\{\psi_0\gamma_5\gamma_i\partial_j\psi_k - \bar{\psi}_i\gamma_5\gamma_0\partial_j\psi_k - \bar{\psi}_i\gamma_5\gamma_j\partial_0\psi_k\}. \tag{30}$$

Clearly the component ψ_0, having no time derivatives, is a Lagrange multiplier, and its coefficient is a constraint

$$\sigma^{ij}\partial_i\psi_j = 0, \tag{31}$$

where γ-algebraic identities have been used in (31). Sudarshan and Johnson and Velo and Zwanziger have shown that (31) has no unique solution (for a review and references, see *e.g.* Deser 1980). One can write $\psi_i = \psi_i^T + \partial_i\chi$, such that

$$\vec{\nabla}\cdot\vec{\psi}^T = 0. \tag{32}$$

This implies, from (31)

$$(\vec{\gamma}\cdot\vec{\nabla})(\vec{\gamma}\cdot\vec{\psi}^T) = 0, \tag{33}$$

which has a solution

$$\vec{\gamma}\cdot\vec{\psi}^T = 0. \tag{34}$$

Equations (32)–(34) reduce the number of physical degrees of freedom of ψ_μ to one, *i.e.*, two helicity states, as is appropriate for a massless particle with spin. The physical component can be projected out,

$$\psi_i^{TT} \equiv \rho_{ij}\gamma_j\gamma_k\rho_{kl}\psi_l, \text{ where } \rho_{ij} = \delta_{ij} - \frac{\partial_i\partial_j}{\nabla^2}. \tag{35}$$

The Rarita–Schwinger action, written entirely in terms of the physical component ψ_i^{TT} is

$$I_{3/2} = -\tfrac{i}{2}\int d^4x\,\bar{\psi}_i^{TT}\not{\partial}\psi_i^{TT}. \tag{36}$$

Now consider pure gravity. It is well known (Arnowitt *et al* 1960ab, 1961) that the Hamiltonian of gravity is a constraint

$$\mathscr{H} = N^\mu R_\mu, \tag{37}$$

where N^μ are Lagrange multipliers, and

$$R^0 \equiv \pi^{ij}\pi_{ij} - \tfrac{1}{2}\pi^2 - ({}^3g)^{1/2}\,{}^3R = 0,$$
$$R^i \equiv -2D_j\pi^{ij} = 0. \tag{38}$$

In the tetrad formalism, we have 16 components for e_μ^a, *i.e.* 6 extra over and above the 10 of the metric. These 6 extra components correspond to the freedom of local Lorentz transformations, and are eliminated as follows:

$$\mathscr{H}\ (\text{tangent space}) = \underset{\substack{\text{[Einstein constr.}\\ \text{in tetrad form]}}}{e_0^aC_a(x)} + \underset{\text{[extra]}}{\tfrac{1}{2}\omega_0^{ab}J_{ab}} \tag{39}$$

Finally, we discuss the case of $N = 1$ supergravity (Teitelboim 1977). We simply add on to (39), the gravitino constraint piece discussed above [(31)]:

$$\mathscr{H}_{SG} = e_0^aC_a(x) + \tfrac{1}{2}\omega_0^{ab}J_{ab}(x) + \bar{\psi}_0S(x). \tag{40}$$

We can verify that the constraints satisfy the algebra

$$
\begin{aligned}
&[J_{ab}(x), J_{cd}(x')] = \eta_{ac} J_{bd}(x)\delta^{(4)}(x-x') + \text{three more terms},\\
&[C_c(x), J_{ab}(x')] = [\eta_{cb} C_d(x) - \eta_{ca} C_b(x)]\delta^4(x-x'),\\
&[C_a(x), C_b(x')] = [\tfrac{1}{2}\Omega_{abcd}(x)\, J^{cd} + {}^*f_{ab} S(x)]\delta^4(x-x'),
\end{aligned}
\tag{41}
$$

where
$$
\begin{aligned}
&\Omega_{abcd} \equiv R_{abcd} - \psi_a \Sigma_{bcd} + \psi_b \Sigma_{acd},\\
&f_{ab} \equiv \tfrac{1}{2}\varepsilon_{abcd}(\partial^c \psi^d - \partial^d \psi^c),\\
&\Sigma_{cab} \equiv \gamma_5(\gamma_c^* f_{ab} + \tfrac{1}{2}\delta_{ac}\gamma^* \cdot f_b - \delta_{bd}\gamma^* \cdot f_a).
\end{aligned}
$$

This is the constraint algebra corresponding to pure gravity. In addition, we have the gravitino constraint algebra

$$
\begin{aligned}
&[S(x), J_{ab}(x')] = -\tfrac{1}{2}\sigma_{ab} S(x)\delta^4(x-x'),\\
&[S(x), C_c(x')] = \Sigma_{cab} J^{ab}(x)\delta^4(x-x'),\\
&\{S(x), S(x')\} = -2\gamma^a C_a(x)\delta^4(x-x').
\end{aligned}
\tag{42}
$$

Suppose S is not known; one could attempt to obtain an expression for S, such that the algebra of constraints is satisfied. This would then lead to the Hamiltonian for supergravity. One makes, following Teitelboim (1977), the 'square root ansatz':

$$
S \equiv \gamma_i \psi_j \pi^{ij} + 4\sigma^{ij} D_i \psi_j, \tag{43}
$$

which is obtained by linearizing the Einstein constraint R^0 in (38) which is quadratic in π^{ij}. Using the fundamental anticommutation relations

$$
\{\psi_i(x), \psi_j(x')\} = -\tfrac{1}{8}\gamma_i\gamma_j\delta^4(x-x'), \tag{44}
$$

and

$$
\{\sigma_{ij} D^i_x \psi^j(x), \sigma_{lm} D^l_{x'} \psi^m(x')\} = \tfrac{1}{64}\, {}^3R(x)\delta^4(x-x'),
$$

one obtains the following remarkable results:

(i) The ansatz (43) satisfies the algebra of constraints
(ii) $S \equiv 0$ has a unique solution, in contrast to the flat space Rarita–Schwinger constraint [(31)].
(iii) We have indeed obtained the supergravity Hamiltonian as a precisely-defined square root (in the sense of Dirac) of the Einstein Hamiltonian for pure gravity.

The constraint analysis presented above has a very interesting by-product. Deser and Teitelboim (1977) have shown that the ADM mass analogue for supergravity is non-negative, as a direct consequence of the square root property. For asymptotically flat spacetimes, the ADM mass is defined as a flux integral over a surface at spatial infinity

$$
P^0 \equiv -\oint_{\Sigma} d\vec{\Sigma} \cdot \vec{\nabla} q^l, \tag{45}
$$

where q^l is the Newtonian component of the spatial metric. We assume that fields fall off sufficiently rapdily as $r \to \infty$. One can, analogously, define a global supercharge

$$
Q \equiv \oint_{\Sigma} d\Sigma^i \sigma_{ij} \psi^j, \tag{46}
$$

assuming $\psi_i(\vec{r}) \xrightarrow[r \to \infty]{} 1/|\vec{r}|^2$

Using the algebra of constraints, these asymptotic generators can be shown to satisfy the same algebra as the flat-space generators, *viz.* the super Poincare algebra [(1)–(3)]. This implies,

$$P^0 = \tfrac{1}{4} \sum_{\alpha=1}^{4} (Q\alpha)^2 \geqslant 0. \tag{47}$$

Thus, the ADM mass for supergravity is positive.

Consider now the following remarkable observation by Grisaru (1978). One can write

$$P^0_{SG} = P^0_1(e) + P^0_2(e, \psi), \tag{48}$$

where

$$P^0_2(e, 0) = 0. \tag{49}$$

We now look at the matrix elements of P^0_{SG} between physical states; these have the functional integral representation

$$\langle | P^0_{SG} | \rangle = \int_{\psi[a]}^{\psi[b]} [\mathrm{d}e][\mathrm{d}\psi] \exp(iS/\hbar)\, P^0(e, \psi). \tag{50}$$

We evaluate the integral for $\psi[a] = 0$ to $\psi[b] = 0$, *i.e.* for zero gravitino asymptotic states. We then restrict ourselves to the classical limit $\hbar \to 0$, which confines us to only tree-diagrams with no gravitino loops. This obviously corresponds to pure Einsteinian gravity. Going back to (47), we deduce, in this limit

$$\lim_{\hbar \to 0} \langle g | P^0_1(e) | g \rangle \geqslant 0, \tag{51}$$

for all physical graviton states. The left side of (51) is the ADM mass for pure gravity, and the inequality is a proof of the positivity of the ADM mass for gravity. It is important to point out that the above proof is by no means rigorous, and has since been improved by Witten (1981) and others. However, the appeal of the above approach is in its simplicity; indeed no detailed solution of constraints was necessary to arrive at (51), but merely the existence of the square root property. Also it appears as though Einstein's gravity 'knows' of its doubly gauge-invariant coupling to the gravitino; the coupling is not actually necessary. 'The marriage need not be consummated!'

6. Superspace

Having established that the supergravity action is the square root of the action for pure gravity, it is logical to enquire as to what the square root of space-time could be. The answer is superspace (Salam and Strathdee 1974)—an augmented version of space-time, where we have added on to the space coordinates x^0, x^1, x^2, x^3, a set of four anticommuting coordinates θ_α, $\alpha = 1, 2, 3, 4$ which are elements of a Grassmann algebra. They obey

$$[x^\mu, \theta^\alpha] = 0 = \{\theta_\alpha, \theta_\beta\} = \{\theta_\alpha, \overline{\theta}_\beta\}. \tag{52}$$

We also introduce the superspace coordinate $z^M = [x^\mu, \theta^\alpha]$. z^M has the following infinitesimal transformation property:

ss transformation $\quad z^M \to (x^\mu + i\bar{\varepsilon}\gamma^\mu\theta, \theta^\alpha + \varepsilon^\alpha)$.

Translation $\quad z^M \to (x^\mu + a^\mu, \theta^\alpha)$.

Homogeneous Lorentz Transformation

$$z^M \to (x^\mu + \tfrac{1}{2}\omega^{\mu\nu}x_\nu, \theta^\alpha + \tfrac{1}{2}\omega^{\mu\nu}\sigma_{\mu\nu}\theta^\alpha). \tag{53}$$

These show that θ^α are Majorana spinorial coordinates. Further, we have the square root property for ss transformations

$$[\delta_1, \delta_2]z^M = (2\bar{\varepsilon}_1\gamma^\mu\varepsilon_2, 0). \tag{54}$$

For extended ss, we have as many θ^α's, as the number of ss generators Q^α's, i.e. θ_i^α, $i = 1, \ldots N$ for N-extended ss. Furthermore, $\{\theta_i^\alpha\}$ transform irreducibly as an N-dimensional representation of the group $O(N)$. Thus, in contrast to the usual spacetime coordinates which remain completely aloof to internal symmetry, the superspace coordinates have components {the θ^α's} that transform *non-trivially* under the action of the internal symmetry group. This is the first time that spacetime (or superspace) geometry has been unified with internal symmetry transformations.

One can now define functions over superspace $V = V(z)$ with definite transformation properties under Poincaré transformations. Under infinitesimal ss transformations, $V(z)$ transforms as

$$[\bar{\varepsilon}Q, V(z)] = \left\{\bar{\varepsilon}\frac{\partial}{\partial\theta} + i\bar{\varepsilon}\gamma^\mu\theta\partial\mu\right\} V(z). \tag{55}$$

Because of the anticommuting nature of the θ^α's, a Taylor expansion of $V(z)$ in powers of θ converges fairly rapidly;

$$V(z) = V(x, \theta) = \sum_{n=1}^{4} \frac{1}{n!} V_{\alpha_1 \ldots \alpha_n}(x)\theta^{\alpha_1} \ldots \theta^{\alpha_n}, \tag{56}$$

where $V_{\alpha_1 \ldots \alpha_n}(x)$ are ordinary fields of various spins belonging to one multiplet. $V(z)$ also includes auxiliary fields. If $V(z)$ is a real Lorentz scalar, it will have fields up to spin 1 (vector). In this case, it can be decomposed non-locally as

$$V(z) = S(z) + T(z) + V(z)^{\text{transverse}} \tag{57}$$

where $S(z)$ and $T(z)$ are chiral (Wess-Zumino type) super multiplets, satisfying differential constraints

$$(1 \pm \gamma_5)\left[\frac{\partial}{\partial\theta^\alpha} + i(\theta\partial)_\alpha\right]\begin{pmatrix} S \\ T \end{pmatrix} = 0. \tag{58}$$

It can be shown that the Wess-Zumino type supermultiplets with Lagrangian given by (6), can be written in superfield notation as

$$\mathcal{L}_S = [S^* S]_{\text{comp } \propto\, \theta^4}. \tag{59}$$

We can now generalize to curved superspace, *i.e.* the superspace of general super-coordinate (z) transformations

$$z^M \to z^M + \xi^M(z) \tag{60}$$

Intrinsic global properties (*i.e.* topology) of superspace are not as yet common

knowledge. For physicists, it is sufficient to focus on the local aspects of superspace. There, one has the choice of the local geometry.

6.1 *Gauge supersymmetry*

Arnowitt and Nath (1978) choose the supermanifold to be locally pseudo-Riemannian, *i.e.* possess a metric structure, but not necessarily of well-defined signature. Further, the tangent space group is taken to be the orthosymplectic group O sp(3,1/4). We have the supertetrad $E^M_A(Z)$, where $A, B, \ldots$ are tangent space indices, and $M, N \ldots$ are world indices. The supermetric is written g_{MN}, and has an analogous relationship to $E^M_A(Z)$ as for ordinary spacetime. The tangent space metric is $\eta_{AB} = \eta_{mn} K C_{\alpha\beta}$, where C is the charge conjugation matrix. The infinitesimal invariant line element in superspace

$$\mathrm{d}S^2 = g_{MN}\,\mathrm{d}z^M \mathrm{d}z^N, \tag{61}$$

with

$$g^{MN} g_{NP} = \delta^M_P \text{ (no torsion).} \tag{62}$$

To define parallel transport, we need to define an affine super connection. By virtue of the pseudo-Riemannian character of superspace, we take this to be the super Christoffel symbol $\Gamma^M_{NP}(g) = \Gamma^M_{PN}(g)$ (symmetry property). Curvature is defined

$$R^D_{ABC} = \partial_A \Gamma^D_{BC} + \Gamma^E_{AC}\Gamma^D_{BE} - (A \leftrightarrow B). \tag{63}$$

The Einstein action is written as

$$I(K) = \int \mathrm{d}z(-g)^{1/2} R, \tag{64}$$

$$\lim_{K \to 0} I(K) \xrightarrow[\text{components}]{} I_{SG} \text{ (usual)} + \text{higher derivative terms.}$$

The field equations are

$$R_{MN} - \tfrac{1}{2} g_{MN} R = 0 \text{ for pure supergravity.} \tag{65}$$

The gauge supersymmetry theory is ultraviolet finite to all orders of perturbation theory, but because of the presence of higher derivative terms, has unphysical particles (ghosts etc.), and therefore cannot correspond to a realistic spectrum.

6.2 *Affine superspace*

Superspace is supposed to have an affine structure to enable definition of parallel transport (Wess and Zumino 1978; Ferrara and van Nieuwenhuizen 1978; Ogievetsky and Sokatchev 1978; Siegel and Gates 1979). The tangent space group is taken to be $O(3,1)$, *viz.*, the homogeneous Lorentz group, and the tangent space is not Riemannian; it has the invariant tensors

$$\eta_{AB} = \begin{pmatrix} \eta_{mn} & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \eta'_{AB} = \begin{pmatrix} 0 & 0 \\ 0 & \gamma^0_{\alpha\beta} \end{pmatrix}$$

which are not inverses of each other. One defines the affine superconnection as

$$\Phi_M = \Phi^{ab}_M J_{ab}, \tag{66}$$

and the super covariant derivative as

$$\mathscr{D}_M \equiv \partial_M - \Phi_M; \quad \mathscr{D}_A = E_A^M \mathscr{D}_M, \tag{67}$$

$$\mathscr{D}_A = (\partial_a, D_\alpha), \text{ where } D_\alpha = (1+\gamma_5)\left[\frac{\partial}{\partial\theta^\alpha} + i(\bar{\theta}\partial\!\!\!/)_\alpha\right]. \tag{68}$$

The curvature and torsion are defined by the equation

$$[\mathscr{D}_A, \mathscr{D}_B\} = -R_{AB}{}^{ab} X_{ab} - C^c_{AB}\mathscr{D}_c, \tag{69}$$

where the commutator is taken between ordinary space-time derivatives, and anticommutator between spinorial derivatives. Bianchi identities for R and C are derived from the super Jacobi identities satisfied by the covariant derivatives:

$$[\mathscr{D}_A, [\mathscr{D}_B, \mathscr{D}_C\}\} + \text{cyclic} = 0. \tag{70}$$

Various formulations exist, corresponding to choice of kinematical constraints on the torsion C. For example

$$C^m_{\alpha\beta} = 2i(\gamma^m)_{\alpha\beta}, \; C^\gamma_{\alpha\beta} = C^p_{mn} = C^p_{\alpha m} = 0. \tag{71}$$

The Bianchi identities then imply that R^{mn}_{AB} and the remaining components of C can be expressed in terms of the super curvature scalar R, the super Einstein tensor $G_{\alpha\dot{\beta}}$ and the super Weyl tensor $W_{\alpha\beta\gamma}$. Constraints on C are equivalent to constraints on $E^A_M(z)$. $R, G_{\alpha\dot{\beta}}, W_{\alpha\beta\gamma}$ satisfy certain differential constraints. The action for supergravity

$$I_{SG} = \int d^4x\, d^4\theta \det E^A_M(x, \theta), \tag{72}$$

reduces to the component supergravity action if, the super torsion constraint is used. In this case $G_{\alpha\dot{\alpha}} \sim (\sigma^a)_{\alpha\dot{\alpha}} Aa(x)$, $R(x, 0) \sim S + iP$, *i.e.* the auxilliary fields for pure supergravity. However, a different set of constraints for $E^M_A(z)$ also lead to the same action for supergravity in terms of component fields.

Superspace is the natural space to discuss supergravity, because of its manifestly supersymmetrically covariant structure. Present research focusses on the solution of the algebra of auxilliary fields in superspace, setting up of a tensor calculus in superspace and also superfield perturbation theory. The global properties of superspace is another area of intense mathematical activity

7. Conclusion

We conclude our survey of supergravity with a few remarks on the ultraviolet properties of the quantized version of the theory. It was shown by 't Hooft and Veltman (1974) that pure source-free gravity has a one-loop finite S-matrix. There exist non-vanishing higher order invariant counter terms that do not vanish when the field equations are used, thus invalidating a finiteness argument at higher orders. Coupled to matter of lower spin (< 2), gravity is infinite even at one loop (Deser *et al* 1974), following the same counterterm arguments.

Pure supergravity (2, 3/2) has better ultraviolet properties than ordinary gravity. A finite S-matrix exists upto two loops (Deser *et al* 1977, Grisaru 1977, Grisaru *et al* 1976). At three loops, one has an invariant non-vanishing counterterm called the

Bel–Robinson tensor, which ushers in once again the same problems as for ordinary gravity (Deser 1980). Two loop finiteness exists even for N-extended supergravity ($N \leqslant 8$), and this is where supergravity has a clear edge over gravity: we have interacting theories of gravity coupled to matter of lower spin (from 3/2 down to 0) which have finite S matrices upto two loops. The case for higher orders is uncertain for general N, and have only been worked out for $N = 2$, where it is not much hopeful. One expects that, with larger N, the enhanced symmetry might lead to more accidental cancellations, but this is merely a hope at present.

We have surveyed a unique gauge theory of gravity with a local fermionic symmetry, providing consistent and rather restrictive couplings of gravity with lower spin matter. Its formal structure is extremely elegant. It also appears to be the natural scenario for unification of all fundamental forces encompassing gravity. There is definite improvement in the ultraviolet behaviour over ordinary gravity. We have not discussed here other areas where supergravity has played a role: better understanding of the cosmological term, explanation of 'accidental' cancellations in quantum gravity and so on. There is scope for more work in the mathematical structure of superspace and also in global quantization schemes to obviate the problems associated with renormalizability. Other aspects, like dimensional reduction, anomalies etc. are also extremely challenging areas.

References

Aragone C and Deser S 1979 *Phys. Lett.* **B86** 161
Arnowitt R, Deser S and Misner C W 1960a *Phys. Rev.* **117** 1595
Arnowitt R, Deser S and Misner C W 1960b *Phys. Rev.* **118** 1100
Arnowitt R, Deser S and Misner C W 1961 *Phys. Rev.* **122** 997
Arnowitt R and Nath P 1978 *Phys. Lett.* **B78** 581
Boulware D, Deser S and Kay J 1979 *Physica (Netherlands)* **A96** 141
Coleman S and Mandula J 1967 *Phys. Rev.* **159** 1251
Deser S 1970 *J. Gen. Rel. Grav.* **1** 9
Deser S 1980 in *Einstein centenary volume* (ed.) A Held (London and New York: Plenum)
Deser S, Kay J H and Stelle K S 1977 *Phys. Rev. Lett.* **38** 527
Deser S and Teitelboim C 1977 *Phys. Rev. Lett.* **39** 249
Deser S, van Nieuwenhwizen P and Tsao H S 1974 *Phys. Rev.* **D10** 3337
Deser S and Zumino B 1976 *Phys. Lett.* **B62** 335
Ellis J 1981 CERN preprint TH.3206-CERN
Fayet P and Ferrara S 1977 *Phys. Rep.* **32** 249
Freedman D Z, van Nieuwenhuizen P and Ferrara S 1976 *Phys. Rev.* **D13** 3214
Ferrara S and van Nieuwenhuizen P 1978 *Phys. Lett.* **B78** 573
Grisaru M 1977 *Phys. Lett.* **B66** 75
Grisaru M 1978 *Phys. Lett.* **B73** 207
Grisaru M, van Niewenhuizen P and Vermaseren J A M 1976 *Phys. Rev. Lett.* **37** 1662
Kibble T W B 1960 *J. Math. Phys.* **2** 212
Langacker P 1981 *Phys. Rep.* **72** 185
Ogievetsky V and Sokatcheve 1978 *Phys. Lett.* **B79** 222
Rarita W and Schwinger J 1941 *Phys. Rev.* **60** 61
Salam̃ A and Strathdee J 1974 *Nucl. Phys.* **B76** 477
Scherk J 1979 in *Mathematical problems in theoretical physics*, Proc. Int. Colloq. (Switzerland: Lausanne)
Siegel W and Gates S J 1979 *Nucl. Phys.* **B147** 77
Teitelboim C 1977 *Phys. Rev. Lett.* **38** 1106
't Hooft G and Veltman M 1974 *Ann. Inst. H. Poincare* **20** 69
van Nieuwenhuizen P 1981 *Phys. Rep.* **68** 189

Wess J and Bagger J 1981 *Lecture notes on supersymmetry*, Princeton
Wess J and Zumino B 1974 *Phys. Lett.* **B49** 52
Wess J and Zumino B 1978 *Phys. Lett* **B79** 394 (and references therein)
Witten E 1981 *Commun. Math. Phys.* **80** 381

Discussion

A. R. Prasanna: (1) What is the significance of the negative sign taken alongwith the Rarita Schwinger Lagrangian in the supergravity action? (2) Would some mathematician comment on the fact that torsion is not significant as it does not give any new topological invariants of the manifold?

P. Majumdar: (1) The negative sign follows from the signature of the metric of Minkowski space (which is $(-1, 1, 1, 1)$ in this case) and also the representation of the γ-matrices, which in this case is the Majorana representation.

C. Sivaram: Regarding long-range fermionic forces, forces with a neutrino-pair exchange have been postulated between fermions (which obeys a $\sim 1/r^5$ law), similar to forces for exchange of a pair of massless spin 3/2 particles. Of course the usual way to avoid long-range effects of massless spin 3/2 particles would be to break the supersymmetry to give them masses.

P. Majumdar: This is correct. Long range $(1/r^5)$ fermionic exchange is avoided by spontaneous breakdown of supersymmetry and the super Higgs phenomenon which gives a small mass to the gravitino.

§ II. GRAVITATIONAL COLLAPSE

INTRODUCTION

A comprehensive treatment of gravitational collapse encompasses a wide variety of features ranging from the purely theoretical considerations to the observational aspects of its astrophysical consequences. A whole array of diverse branches of physics, such as general relativity, nuclear physics and elementary particle theory, as well as advanced computational techniques have to be employed in order to describe the processes inherent to this phenomenon. In the following articles, we shall sample some of the ingredients that constitute the collapse scenario.

In building up the picture of gravitational collapse, a pivotal role is played by the equation of state of matter at high densities. Significant progress has been made in obtaining realistic equations of state by taking into consideration nuclear interactions as well as the resulting emission of neutrinos. Once the equation of state is given, general relativity steps in to depict the dynamics of collapse through the Einstein field equations. One is faced with a multitude of complexities here—rotation, non-sphericity, magnetic fields, gravitational radiation and so on. Not only the most fundamental issues like the cosmic censorship conjecture are involved in these calculations, but also concrete problems of detail figure prominently. One such question is whether in the course of collapse an appreciable fraction of the collapsing matter can be ejected thereby triggering a supernova explosion. This is where theoretical computations can make important contribution to the astrophysical investigation of the types of stars that might give rise to supernovae.

If the end product of the collapse is a neutron star rather than a black hole, then the allowed equilibrium configurations and their mass-radius distribution become an important element. General relativistic models can be built for such a study on the most general grounds not only utilizing our latest knowledge of equations of state, but also making allowance for our ignorance above and beyond the limiting densities below which these equations of state are valid.

Finally, purely general relativistic effects of the highly curved spacetime geometry engendered by the collapsing matter on zero rest-mass particles, such as neutrinos, can be analysed in detail. The situation is idealized by ignoring other interactions between the neutrinos and the matter.

It is hoped that the next few articles will offer a glimpse of one of the most complex and fascinating areas of gravitational physics.

Equations of state, neutrinos and supernova explosions*

B BANERJEE

Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005, India

Current stellar evolution calculations show that massive stars develop central cores of $1.5\,M_0$ consisting of iron-peak elements and supported by electron degeneracy pressure. As more of the heavier nuclei disintegrate to form elements around ^{56}Fe, the core slowly increases in mass and density; so does the electron chemical potential, making it possible for them to be captured by nuclei. This leads to the gradual 'neutronization' of the nuclei and lowering of the electron density. The subsequent behaviour of this supernova matter, consisting of nuclei, nucleons and leptons depends on its equation of state (EOS). The recent progress in the theory of the EOS is discussed as also the role of the neutrinos, which are copiously emitted during the capture of electron by the nuclei and nucleons. The significance of the change of the adiabatic index of the supernova matter with temperature and density for the supernova explosion is also pointed out.

References

Arnett W D 1979 in *Sources of gravitational radiation* (ed.) L Smarr (Cambridge: University Press)

Lattimer 1981 *Annu. Rev. Nucl. Part. Sci.* **31** 337

Rees M J and Stoneham R J 1982 *Supernovae: a survey of current research* (Dordrecht: O Reidel Publishing Company)

Wheeler J C 1981 *Rep. Prog. Phys.* **44** 85

Discussion

R. Pratap: If Bubble is a phase transition, will the entropy concept be still valid?

B. Banerjee: Like in other kinds of phase transitions the concept of entropy should be valid here also.

S. B. Khadkikar: How sensitive are the calculations to the set of parameters used say SIII or SV (without density dependence)?

B. Banerjee: It should not be very sensitive. But we have not done the calculations with any other set.

K. Kar: For the neutrino trapping problem, Bethe *et al* constructed a diffusion equation and found the diffusion time to be larger than the time of the contraction. Did you have some mechanism to calculate the same in your calculations?

*Summary of the talk

B. Banerjee: We have not started looking into the hydrodynamics so far. We have been concentrating on the equation of state.

K. Kar: When would the effects of pion condensation set in?

B. Banerjee: Almost all currently available calculations find that the critical density for pion condensation is somewhat above the nuclear density. Such densities are probably not reached in the supernova problem.

Harish Bhatt: Does the shortening of the neutrino mean free path make any change in the energy of the supernova ejecta?

B. Banerjee: The energy of the ejecta depends on the strength of the shock wave (that blows off the mantle), which in turn depends on the mean free path of the neutrinos.

M. N. Rao: What is the role of nuclear statistical equilibrium and which is the limit for this reaction?

B. Banerjee: A typical nucleus of A = 500, 1000 is assumed and the surfaces of minimum energy configuration etc are calculated.

Alok Ray: Precisely how is the temperature dependence put in, in the equation of state? Presumably to go from the Skyrme interaction to your temperature dependent equation of state you need to make assumptions about the density of excited states etc. How do you get for example the surface energy of the nucleus (when it is T-dependent)?

B. Banerjee: The temperature dependence enters only through the density ρ_q and kinetic energy density τ_q.

General relativity and gravitational collapse*

C V VISHVESHWARA
Raman Research Institute, Bangalore 560 080

Gravitational collapse is one of the few natural phenomena that require the general theory of relativity for their proper description. The study of gravitational collapse began in the thirties. First came the mass limits for stars or stellar cores that had exhausted their nucleal fuel. The pioneering work of Chandrasekhar gave the mass limit of about $1{\cdot}4\, M_{\odot}$ below which stars could remain in equilibrium as white dwarfs with the degeneracy pressure of electrons balancing gravitation. Likewise Oppenheimer and Volkoff obtained the mass limit of $0{\cdot}7\, M_{\odot}$ for neutron stars in which the degenerate neutrons that constitute the star provide the necessary pressure. More recently this limit has been pushed upto 3–$5\, M_{\odot}$ on the most general grounds. Any stellar core more massive than this is believed to undergo catastrophic collapse resulting in a black hole (Miller and Sciama 1980).

The first and the simplest picture of gravitational collapse was presented by Oppenheimer and Snyder. The model consists of spherically symmetric, homogeneous pressure-free dust contracting under its own gravity. The interior is represented by the Friedmann cosmological metric which is matched on at the contracting surface to the Schwarzschild exterior spacetime. As the pressure is zero, the configuration inevitably collapses to a black hole irrespective of the mass of the sphere. This is the only analytical description of collapse that has been obtained so far. Essentially all work that has been done subsequently has employed numerical computation. The Oppenheimer–Snyder collapse represents a highly idealized situation. More realistic approaches depart from it mainly in three characteristic features: (i) spherical collapse with non-zero pressure, (ii) inclusion of rotation, (iii) non-spherical collapse without rotation. An important by-product of nonspherical collapse is the generation of gravitational radiation. Some typical papers incorporating these aspects were discussed in order to give a broad idea of the state of the art in this field.

There have been several papers to date that consider in detail the spherical collapse with non-zero pressure. One of the more recent ones among them is by Shapiro and Teukolsky (1980) which brings out clearly the general relativistic aspects of the problem. The mathematical technique employed involves the Arnowitt–Deser–Misner (ADM) formalism. Maximal time slicing is used instead of comoving or cosmic time. This helps in "postponing" the appearance of singularities thereby enabling one to follow the formation of the black hole. Depending on the combination of the equation of state, the initial condition and the total mass of the sphere, three distinct possible processes were observed. These are: (i) Homologous core bounce in which about 50 % of the inner core undergoes homologous bounce sending a shock wave through the outer in-falling

*Summary of the talk

matter. This offers the possibility of ejection of the outer layers leading perhaps to a supernova explosion. (ii) Nonhomologous bounce, in which the shock emanating from the central region heats up the entire core which comes to a quasistatic equilibrium corresponding to an equation of state different from the original one. Dissipation of heat through neutrino transport would ultimately initiate further collapse to a black hole. No mass ejection is possible during this process. (iii) Dynamic collapse to a black hole without mass ejection. In this case, the geometry of the interior spacetime was "probed" by tracing the outgoing null geodesics revealing the formation and evolution of the event—and the apparent horizons.

The inclusion of rotation into the scenario of collapse is still in a very rudimentary stage. One of the most important issues here is the validity of the cosmic censorship conjecture: whether the end product will always be a Kerr black hole or whether a naked singularity can appear in some cases. The Kyoto group has been engaged in systematically developing the basic analytic tools *via* ADM formalism towards a numerical description of axisymmetric collapse with rotation (Nakamura 1981). The computational results published by the group seem to be rather preliminary in nature. Calculations have been carried out varying four different factors, namely the equation of state, the initial angular velocity profile, the mass, and the angular momentum divided by the square of the mass. Depending on the values of the last two parameters event horizons are found to form in some cases. In no case was a naked singularity observed. As the computations were not carried out to completion, these findings are hardly conclusive especially in shedding light on the cosmic censorship conjecture. A new feature that emerged in these calculations was the generation of highly energetic jets which might be of astrophysical relevance.

An example of non-spherical collapse without rotation has been presented by Smarr *et al* (1980). The process involves complicated fluid flow exhibiting bounces in both the polar and equatorial directions. A novel feature arising here is the appearance of circulations left in the wake of shocks which may play a significant role in mass ejection.

As has been mentioned earlier, an essential part of non-spherical collapse is the accompanying gravitational radiation. This aspect has been treated in two methods of approximation. In the first method, the collapse itself is described within the framework of Newtonian theory and the gravitational radiation computed in the weak field approximation. For instance, Saenz and Shapiro (1978) consider the collapse of cores that are non-spherical due to rotation or due to internal magnetic fields. Different quantities pertaining to the gravitational radiation, such as the power spectrum, efficiency and the production rate of the radiant energy are computed. These quantities as well as the form of the wave trains reflect some of the dominant features of the collapse dynamics. The second approach is fully general relativistic, but is confined to perturbations superposed on spherical collapse. In a series of papers Cunningham *et al* (1978, 1979, 1980) have considered such perturbations in the first and second orders. The production and the characteristics of the resulting gravitational radiation are found to be largely dependent on the exterior Schwarzschild spacetime. For instance a burst of radiation is produced when the surface approximately crosses the radius $r = 3m$. This is followed by damped oscillations corresponding to the complex quasi-normal modes admitted by the exterior spacetime and finally by a decaying tail with the time dependence of $t^{-(2l+2)}$. It is not at all certain how many of such characteristics would be retained in the non-perturbative case of strongly distorted collapse.

Apart from the task of putting together a composite picture of realistic gravitational

collapse, many problems remain at individual levels. Even to begin with one does not know the fully relativistic rotating equilibrium configurations that would constitute the pre-collapse models. Further, the inclusion of magnetic fields within the collapse dynamics has not even been attempted yet. Nor has gravitational radiation in a highly non-spherical collapse taken into consideration without approximation. When all these and other problems have been handled properly, theoretical calculations can be hoped to yield valuable information pertinent to astronomical observations.

References

Cunningham C T, Price R H and Moncrief M 1978 *Astrophys. J.* **224** 643
Cunningham C T, Price R H and Moncrief M 1979 *Astrophys. J.* **230** 870
Cunningham C T, Price R H and Moncrief M 1980 *Astrophys. J.* **236** 674
Miller J C and Sciama D W 1980 in *General relativity and gravitation* (ed.) A Held (New York: Plenum Press) vol. 2
Nakamura T 1981 in *Evolution of the universe, stars and the solar system* (Suppl. to Progr. Theor. Phys., Jpn **70** 202)
Saenz R A and Shapiro S L 1978 *Astrophys. J.* **221** 286
Shapiro S L and Teukolsky 1980 *Astrophys. J.* **235** 199
Smarr L, Taubes C and Wilson J R 1980 *Essays in general relativity* (ed.) F J Tipler (New York: Academic Press)

Core envelope models of collapsed objects

B R IYER

Raman Research Institute, Bangalore 560080, India

1. Introduction

It is well known that in the framework of general relativity (GR) and other relativistic gravitational theories neutron stars (NS) have a maximum rest mass. This has led to considerable interest* in setting up maximum mass limits because this value plays a central role in distinguishing between a neutron star and a black hole (BH).

No information on the maximum mass can be obtained from GR alone. Some information about the properties of matter is essential and the more stringent this requirement, the more constrained will be the mass limit. The calculation of the range of NS mass would be a routine matter if the equation of state (EOS) of matter at the end of thermonuclear evolution were known at high densities. Since this is not so, the best we can do is to avoid unnecessary extrapolations of the EOS and use instead a minimum number of general assumptions of the NS matter. This was done by Rhoades and Ruffini (1974) who obtained a mass limit of 3·02 $M_\odot$. This was followed by a more systematic investigation by Sabbadini and Hartle (1973) who relaxed the general requirement of 'causality' and obtained a mass limit of 5 $M_\odot$. Later Chitre and Hartle (1976) established the 3 $M_\odot$ limit more rigorously by restoring the 'causality' requirement.

The feature we critically looked into initially within the framework of these models was the question of size or radii of these objects (Dhurandhar and Vishveshwara 1981). As is well known, the general relativistic effects of a distribution of mass is determined not by its mass alone but by its radius *i.e.* the degree of its compactness. General Relativistic effects are determined by the ratio of $R:M$. For a BH, a totally relativistic object this ratio is 2 while objects with $R:M$ greater than 6 are almost Newtonian. It is between these values that objects show strange and interesting properties and for convenience we subdivide them into two classes.

(a) compact objects for which $3 < R/M < 6$,

(b) ultra compact objects for which $2 < R/M < 3$.

For ultracompact objects the general relativistic effects are stronger and lead to unusual properties *e.g.*,

(i) For photons (exactly) and neutrinos (approximately) studies show that the exterior Schwarzschild potential has a barrier at $r = 3M$. Consequently the behaviour of these zero mass particles is different since there is a possibility that some of them may be trapped by such ultracompact objects. In fact our original interest in such objects

*This article draws heavily on Hartle (1978). See also Chandrasekhar (1931), Oppenheimer and Volkoff (1939).

came from related studies on neutrinos in compact objects and gravitational collapse (Dhurandhar 1983).

(ii) The analysis of the tail of x-ray bursts shows that it can be best fitted by a black body at temperature T. The luminosity L is related to the observed flux F and distance d by

$$L = 4\pi d^2 F. \tag{1}$$

In terms of the temperature T

$$L = 4\pi R^2 \sigma T^4, \tag{2}$$

where R is the observed black body radius and σ the Stefan's constant. Thus

$$R = \sigma^{-1/2} F^{1/2} T^{-2} d. \tag{3}$$

The effect of the gravitational redshift is to decrease the observed values of F and T relative to the values F_0 and T_0 on the NS surface, *i.e.*

$$F = F_0(1+Z)^{-2}, \; T = T_0(1+Z)^{-1}, \tag{4}$$

so that $\quad R = (\sigma^{-1/2} F_0^{1/2} T_0^{-2} d)(1+Z),$

i.e. $\quad R = R_0(1+Z) = R_0\left(1 - \dfrac{2M}{R_0}\right)^{-1/2}, \qquad (5)$

where R_0 is the real black body radius of the NS. For $R_0 < 3M$ as mentioned above only a fraction $\sin^2\delta$ of the emitted flux escapes where

$$\sin\delta = 3\sqrt{3}\frac{M}{R_0}\left(1 - \frac{2M}{R_0}\right)^{1/2}, \tag{6}$$

so that F is multiplied by this additional factor leading to

$$R = 3\sqrt{3}\,M. \tag{7}$$

This analysis of Van Paradijs (1979) shows that the real and observed black body radius are related in a different manner for compact objects as compared to ultracompact objects. Thus for a fixed M the observed R has a minimum value of $3\sqrt{3}\,M$ for $R_0 < 3\,M$. Also for a particular observed value of black body radius, the mass of the NS cannot be larger than

$$M = R/3\sqrt{3} = \frac{R\,(\text{km})}{7{\cdot}7}\,M_\odot. \tag{8}$$

This seems to offer a method of putting constraints on the M–R relation observationally modulo the identification of black body radius with the physical radius.

Our preliminary interest is to look for such ultracompact objects within the framework of Sabbadini–Hartle–Chitre models which for reasons that will be clear are quite general.

2. The core envelope model

The minimal general assumptions made in this model are:

(A1) Matter is a perfect fluid described by a one-parameter EOS relating pressure p and energy density ρ

$$p = p(\rho). \tag{9a}$$

(A2) The energy density ρ is non-negative

$$\rho \geqslant 0. \tag{9b}$$

(A3) Matter is microscopically stable

$$(\mathrm{d}p/\mathrm{d}\rho) \geqslant 0, \; p \geqslant 0. \tag{9c}$$

(A4) The EOS is known definitely below a fiducial density ρ_0.

The above minimal restrictions on the properties of NS matter lead to a number of useful deductions on the structure of spherical NS in GR. The spacetime geometry of a spherical NS is described by the metric which in Schwarzschild coordinates are given by

$$\mathrm{d}s^2 = -\exp[\nu(r)]\,\mathrm{d}t^2 + \exp[\lambda(r)]\,\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2). \tag{10}$$

The Einstein field equations

$$G^\mu_\nu = 8\pi T^\mu_\nu, \tag{11}$$

and the equation of hydrostatic equilibrium

$$T^\mu_{\nu;\mu} = 0, \tag{12}$$

can be rewritten giving the Tolman–Oppenheimer–Volkoff (TOV) equations

$$\mathrm{d}m/\mathrm{d}r = 4\pi r^2\rho, \tag{13a}$$

$$\mathrm{d}p/\mathrm{d}r = -\frac{(\rho+p)(m+4\pi r^3 p)}{r^2(1-2m/r)}, \tag{13b}$$

$$\mathrm{d}\nu/\mathrm{d}r = \frac{2(m+4\pi r^3 p)}{r^2(1-2m/r)}, \tag{13c}$$

$$\exp(-\lambda) = \left(1 - \frac{2m(r)}{r}\right). \tag{13d}$$

To obtain a stellar model these equations must be integrated from the centre $r = 0$ with central density ρ_c to the surface $r = R$ where pressure vanishes. The three boundary conditions (BC) are

$$p(r=0) = p(\rho_c) = p_c, \tag{14a}$$

$$m(0) = 0, \tag{14b}$$

$$\exp[\nu(R)] = 1 - \frac{2m(R)}{R} = 1 - \frac{2M}{R}, \tag{14c}$$

where R and M are the total radius and mass of the object. Assumptions A1–A3 and the structure equations (13a) and (13b) imply

(L1) No region of a spherical symmetric star can be inside its gravitational radius

$$\frac{2m(r)}{r} < 1 \tag{15}$$

(L2) The density is non-increasing outwards

$$(\mathrm{d}\rho/\mathrm{d}r) \leqslant 0. \tag{16}$$

(L3) The average density is also non increasing outwards

$$\frac{\mathrm{d}}{\mathrm{d}r}(m/r^3) \leqslant 0. \tag{17}$$

Since the density is non-increasing outwards the radius at which the density assumes the fiducial value ρ_0, divides the star into two parts.

(i) An envelope $r \geqslant r_0$, $\rho < \rho_0$ where the EOS is known

(ii) A core $r < r_0$, $\rho \geqslant \rho_0$ where the NS matter only satisfies the general requirements (A1–A4). The BC required to integrate (13a) and (13b) are now

$$m(r_0) = M_0, \quad p(r_0) = p(\rho_0) = p_0. \tag{18}$$

The total mass of the star is

$$M = M_0 + M_e(r_0, M_0), \tag{19}$$

where M_e is a computable function from the assumed EOS for $\rho < \rho_0$.

To proceed further one should be able to obtain the allowed range of core masses and radii *i.e.* the allowed region in the $r_0 - M_0$ plane. The optimum upper bound on the total mass is then found by maximizing (19) over the minimal allowed range of variables r_0 and M_0.

3. Core limits

(i) The non-increasing density implies that its lowest value in the core is at its boundary. Thus

$$\mathrm{d}\rho/\mathrm{d}r \leqslant 0 \Rightarrow M_0 \geqslant \tfrac{4}{3}\pi r_0^3 \rho_0. \tag{20}$$

This gives the lower boundary—a cubic in the $r_0 - M_0$ plane.

(ii) To find the upper boundary of the allowed region we extend the method developed by Buchdahl (1959) to obtain bounds on redshifts of stars: Let

$$\alpha = \exp[\nu(r)/2)]. \tag{21}$$

Using (13c) we obtain

$$\mathrm{d}\alpha/\mathrm{d}r = \frac{m + 4\pi r^3 p}{r^2(1 - 2m/r)}\alpha. \tag{22}$$

Employing the structure equation to eliminate p, (22) yields

$$\mathrm{d}^2\alpha/\mathrm{d}\xi^2 = \frac{\alpha}{r}\frac{\mathrm{d}}{\mathrm{d}r}(m/r^3) \tag{23}$$

where $$\mathrm{d}\xi/\mathrm{d}r \equiv \frac{r}{(1 - 2m/r)^{1/2}}. \tag{24}$$

From (17) the right side is non-positive, equalling zero for a uniform density star. Hence

$$\mathrm{d}^2\alpha/\mathrm{d}\xi^2 \leqslant 0. \tag{25}$$

For such a convex downward curve the slope at any point is always less than that of the chord joining $\alpha(\xi)$ and $\alpha(0)$ and since $\alpha(0) \geqslant 0$ it follows that

$$\frac{1}{\alpha}\frac{\mathrm{d}\alpha}{\mathrm{d}\xi} \leqslant \frac{1}{\xi}. \tag{26}$$

The right side is maximum for a uniform density star and using (17) after some algebra it follows that

$$\frac{1}{\xi} \leqslant \frac{2m/r^3}{[1-(1-2m/r)^{1/2}]}. \tag{27}$$

From (26), (27) and (13c) it follows that

$$M \leqslant \tfrac{2}{9}r[1-6\pi r^2 p+(1+6\pi r^2 p)^{1/2}]. \tag{28}$$

When (28) is evaluated for a NS surface where $p = 0$ it gives

$$2M/R \leqslant 8/9, \tag{29}$$

which is the origin of the surface redshift bound of $Z_s \leqslant 2$.

At the boundary of the core $p = p_0$ and (28) yields

$$M_0 \leqslant \tfrac{2}{9}r_0[1-6\pi r_0^2 p_0+(1+6\pi r_0^2 p_0)^{1/2}], \tag{30}$$

where the equality sign holds for a uniform density star with infinite central pressure. Hence this bound cannot be improved and for models where $p_0 \ll \rho_0$ it becomes linear

$$M_0 \leqslant \tfrac{4}{9}r_0. \tag{31}$$

The core with the largest mass is constructed from matter with the stiffest equation of state—incompressible constant density matter. There is a limit on mass that can be contained within r_0 without pressure becoming infinite at the centre and destroying the equilibrium. This is the physical origin of the upper bound. The allowed region is shown in figure 1.

In the Sabbadini–Hartle model matter in the envelope is assumed to satisfy the Baym–Bethe–Pethick–Sutherland (BBPS) EOS which is the most accurate at present for cold neutron matter (Baym *et al* 1971, 1972). Unlike earlier equations it takes into account consistently the contribution of lattice energy to pressure (not very significant) as well as in determining the equilibrium nuclides at a particular density. With this Sabbadini and Hartle obtained a limit of $5M_\odot$. As mentioned earlier any further assumption on the matter content will induce a more stringent limit. Chitre and Hartle (1976) put in the additional constraint that

(A5) $$(\mathrm{d}p/\mathrm{d}\rho)^{1/2} \leqslant 1 \tag{32}$$

Physically $(\mathrm{d}p/\mathrm{d}\rho)^{1/2}$) is the hydrodynamic phase velocity of sound waves in NS matter. In the absence of dispersion and absorption this would be the velocity of signals in the medium and (32) would be the condition that signal velocity be less than velocity of light i.e. matter is 'causal'. However NS matter is dispersive and from general consideration of causality there seems no reason to require (32). Whether it is true in general for matter at high densities remains to be demonstrated though most calculations on EOS at high densities satisfy it.

Buchdahl's method cannot be extended in any straightforward way to take care of

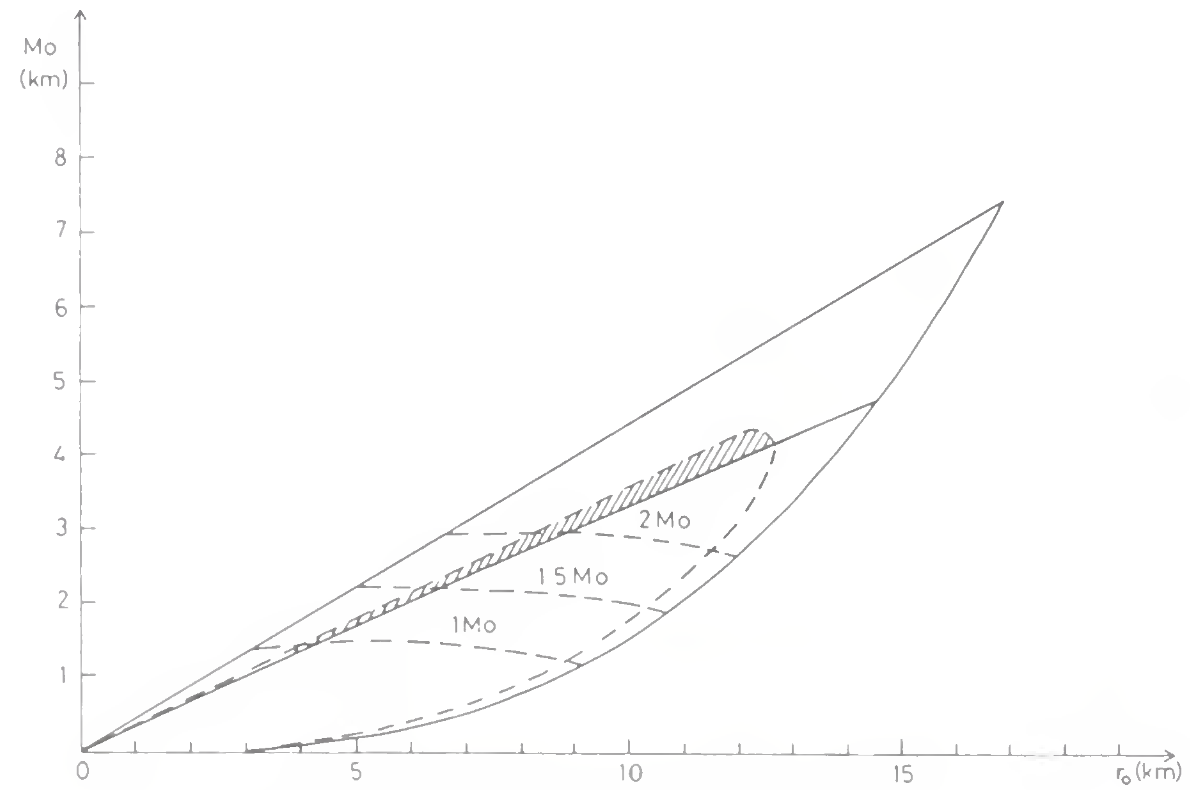

Figure 1. The figure shows the maximal region (drawn with broad unbroken line) of permitted cores in the (r_0, M_0) plane. The inner broad unbroken line depicts the region occupied by causal cores. The thin unbroken line corresponds to $r = 3m$ objects and divides the cores into two sections corresponding to $r < 3m$ and $r > 3m$ objects, the region above the line corresponding to $r < 3m$ objects. The shaded region is the intersection of the regions corresponding to causal cores and to $r < 3m$ objects. The thin dashed curves correspond to constant mass objects namely 1, 1·5 and 2 solar mass.

this constraint and it is found that treating the problem as a variational problem with constraints given by the assumptions A1–A5 and the structure equations (13a, b) the allowed region in the r_0–M_0 plane can be obtained. We skip the detailed derivation but quote the final result, that configurations that extremize the core mass at a given radius are those for which density is piecewise continuous with $\mathrm{d}p/\mathrm{d}\rho = 1$ except at density discontinuities. Numerical computations give the allowed region shown in figure 1 and coincides with a minimal region where $p = \rho - \rho_0 + p_0$ and no density discontinuity in the core or at the boundary. The maximum mass can be obtained and it turns out to be $3\,M_\odot$.

4. Integration of equations and discussions

Equations (13) were integrated by a 4-point Runga–Kutta method with variable step length since the derivatives of the functions have a large range of values. The allowed range was covered by a grid at 25 points, each grid value giving one set of b.c., equation (18). The fiducial density was chosen to be

$$\rho_0 = 5{\cdot}09 \times 10^{14}\ \mathrm{g/cc}. \tag{33a}$$

and the corresponding pressure is

$$p_0 = 7{\cdot}39 \times 10^{33}\ \mathrm{dynes/cm^2}. \tag{33b}$$

This value of the fiducial density is chosen since it represents the highest value to which BBPS believe their nuclear matter calculation to be applicable. Equations (13) are

integrated till p drops to 0 and this yields the values of the NS radius and mass. To trace the $R/M = 3$ curve the following observation proves useful. For a fixed core radius r_0 the mass increases with increasing core mass. Hence by the standard root finding methods for a fixed r_0 one moves in the allowed region till R/M becomes 3. This curve is shown in figure 1. It starts with a slope 1/3 and bends back to cut the lower boundary at $r_0 = 14{\cdot}4$ km. The points above this curve correspond to $r < 3M$ objects and they make up about half the permitted region. Moreover there is a non empty intersection with the region of causal cores.

Finally the dashed lines almost transverse to the maximal region represent stars with constant masses. It is interesting *e.g.* that for $1{\cdot}5\, M_\odot$, 20 % are ultracompact and even if only causal cores are allowed there is still 8 % of such objects. It should be mentioned that these curves are in effect curves of the distribution of GR effects and the above numbers could be interesting in problems where GR energy shifts are crucial.

The results of table 1 can be qualitatively understood as follows. For the BBPS EOS $p \ll \rho$ everywhere. Hence

$$\mathrm{d}p/\mathrm{d}r = -\rho f(r, m), \tag{34}$$

Table 1. Values of total radius R, total mass M and their ratio $R{:}M$ for different values of core radius r_0 and mass M_0.

r_0(km)	m_0(km)	R(km)	M(km)	$R{:}M$
	0·237	4·386	0·261	16·799
	0·439	3·389	0·448	7·564
2·809	0·642	3·098	0·646	4·797
	0·884	2·965	0·846	3·504
	1·046	2·887	1·047	2·757
	0·647	7·283	0·752	9·681
	1·017	6·469	1·070	6·049
5·618	1·387	6·099	1·416	4·306
	1·757	5·892	1·774	3·322
	2·127	5·763	2·136	2·698
	1·404	9·760	1·594	6·124
	1·873	9·268	1·990	4·658
8·427	2·341	8·964	2·415	3·712
	2·809	8·762	2·855	3·069
	3·277	8·616	3·303	2·608
	2·682	12·162	2·914	4·174
	3·144	11·912	3·311	3·579
11·236	3·606	11·717	3·725	3·146
	4·069	11·565	4·150	2·787
	4·531	11·443	4·582	2·497
	4·653	14·574	4·857	3·000
	4·970	14·475	5·137	2·818
14·044	5·288	14·388	5·421	2·654
	5·606	14·311	5·708	2·507
	5·924	14·240	5·999	2·374

where

$$f(r, m) = m/r(r-2m). \tag{35}$$

The following features follow

(i) f is an increasing functon of m but a decreasing function of r. For a star with a large value of core mass the pressure gradient is large and the envelope is thinner and conversely. Similarly for a small value of core radius the gradient being larger the envelope is thin.

(ii) On the upper boundary $m \sim \frac{4}{9} r_0$; $f \sim 4/r_0$. As $r_0 \to 0$ f is infinitely large and the envelope is a thin shell. On the lower boundary $M_0 = \frac{4}{3}\pi r_0^3 \rho_0$; $f \sim \frac{4}{3}\pi r_0 \rho_0$ so that as $r_0 \to 0$, $f \to 0$. Hence the star radius is large. It can be shown that in the limit the envelope becomes the entire star it has the same structure as a non-rotating star with central pressure determined by the slope of the line in the r_0–M_0 plane.

(iii) The upper bound on the mass occurs at the maximum value of the core mass in the allowed region with the contribution of the envelope smaller than 1%. Around our chosen ρ_0 the envelope gives an unimportant contribution to the bound so that it does not matter which of the several not too different EOS are used. This does not apply to higher value of ρ_0 in general. The outermost crust contributes negligibly to total mass since the density here is very low.

(iv) There are indications that the radius of the NS is sensitive to the EOS. Consequently with a different EOS, the $R = 3M$ curve will be shifted and hence the percentage of objects with causal cores could crucially depend on the EOS.

We thus see that within the core envelope models ultracompact objects do exist. However, we feel that the existence of such objects is of no real consequence unless one can say something about the stability of these configurations. Unless proved otherwise the question of whether the additional requirement of stability will lower the maximum mass limit is a fascinating question. Note that there is a slight difference in that this is not a general requirement on the EOS but on the configuration. Is it only a coincidence that the masses of the NS established through studies of binary x-ray and radio pulsars are constrained to the range 1·2–1·6 $M_\odot$ (Kelley and Rappaport 1980).

5. Stability

To investigate this question one initially addresses oneself to the question whether such objects are stable under radial perturbations. The detailed equations have been set up by Chandrasekhar (1964) and we quote them.

Let the adiabatic motion of the star in its nth normal mode be given by an amplitude.

$$\delta(r, t) = \frac{\rho^{\nu/2}}{r^2} u_n(r) \exp(i\omega_n t). \tag{36}$$

The eigenequation for u_n has the Sturm Liouville form

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(P\frac{\mathrm{d}u_n}{\mathrm{d}r}\right) + (Q + \omega_n^2 W)u_n = 0, \tag{37}$$

where P, Q, W are known functions of γ, ν, λ, p, ρ.

$$P = \gamma p r^{-2} \exp(\lambda + 3\nu)/2, \tag{38a}$$

$$Q = -4\frac{dp}{dr}r^{-3}\exp(\lambda+3\nu)/2 - 8\pi p(\rho+p)r^{-2}\exp 3(\lambda+\nu)/2 \tag{38b}$$

$$+ (dp/dr)^2(\rho+p)^{-1}r^{-2}\exp(\lambda+3\nu)/2,$$

$$W = (\rho+p)r^{-2}\exp(3\lambda+\nu)/2, \tag{38c}$$

$$\gamma = (\rho+p)/p\,(\partial p/\partial p)_{\text{EOS}}. \tag{38d}$$

Acceptable solutions to (37) must satisfy boundary conditions at the centre and the stellar surface. Since the origin is a regular singular point the finiteness of δr and $d/dr\,(\delta r)$ at the origin imply $u_n \sim r^3$ there. Physically it means that the fluid at the centre is not set into radial oscillation. At the surface (which is pulsating) the physical condition is that the Lagrangian change in pressure vanish.

$$\Delta p = -\exp(\nu/2)(\gamma p/r^2)\frac{du_n}{dr} = 0. \tag{39}$$

For models of our type since the other variables are finite this is equivalent to

$$du_n/dr = 0 \quad \text{at} \quad r = R. \tag{40}$$

Instability against radial perturbations manifest as complex eigenfrequencies *i.e.* negative ω_n^2. Tests of stability are essentially methods of determining existence of modes with negative ω_n^2 and these methods are catalogued by Bardeen *et al* (1966). We follow the method originally developed by Bardeen (1965) which depends essentially on some general properties of the Sturm Liouville problem. These are

(a) There exist a discrete spectrum of eigen frequencies.

(b) The nth normal mode has n nodes between the two boundary points.

The method consists of the following. Set $\omega_n^2 = 0$ and integrate from the surface where $u_n \neq 0$ $u_n' = 0$ inwards upto the centre. If by this point the function has gone through N nodes then the modes $n = 0, 1, N-1$ are unstable while those above are stable. It should be noted that even if there are many unstable modes the lowest one will still dominate. The configuration will be stable only if none of the normal modes has negative ω_n^2, in other words, the fundamental mode itself is stable. Thus integrating inwards with $\omega_n^2 = 0$ if the function does not cut the $u = 0$ axis the configuration is stable.

6. Results

The method outlined above was first applied to the core envelope models constructed with the BBPS EOS in the envelope. This was done primarily to investigate whether the ultracompact objects found in the previous sections within these models remain or whether they are all unstable. Later the complete grid of 25 points in the allowed region was scanned. Since one needs the EOS to check the stability, within the core envelope models the stability can be checked only for the envelope and this was done at first. The result was the following: When only the envelope is considered no object shows an instability *i.e.*, all are stable under radial perturbations.

To proceed further one should be able to say something about the EOS in the core. As mentioned earlier though in the general case no EOS is put in the core when the assumption of 'causal' cores is invoked one is automatically led to the fact that

Table 2. Mass and radius of a neutron star with a limiting causal core and a BBPS envelope. All configurations below $\rho_c = 1{\cdot}13E-3$ are stable.

$\rho_c(\text{km})^{-2}$	$M(M_0)$	R(km)	R:M
4·0 E-4	0·636	10·73	11·440
4·5 E-4	1·556	12·21	5·317
5·0 E-4	2·080	12·86	4·191
6·0 E-4	2·588	13·26	3·472
7·0 E-4	2·806	13·28	3·208
8·0 E-4	2·911	13·20	3·074
9·0 E-4	2·961	13·08	2·995
1·0 E-3	2·984	12·96	2·943
1·13E-3	2·992	12·80	2·900
2·0 E-3	2·889	12·01	2·819
4·0 E-3	2·669	11·19	2·843
6·0 E-3	2·545	10·83	2·885
8·0 E-3	2·466	10·63	2·922
1·0 E-2	2·413	10·51	2·953

stationary configurations in the limit satisfy the EOS

$$p - p_0 = \rho - \rho_0. \tag{41}$$

The stability of configurations with the above EOS in the core and the BBPS EOS in the envelope was then tested for by the method described in the previous section. This was done because as is well known if M is the total mass and ρ_c the central density, then $\mathrm{d}M/\mathrm{d}\rho_c > 0$ is only a necessary but not sufficient condition for stability. There is a possibility that at the maximum the configuration changes not from stability to instability but from one mode instability to two-mode instability as in the Misner Zapolsky (1964) case. The results of the numerical computation are given in table 2 and it turns out that taking into account the jump in u_n' at the discontinuity in γ, the situation is indeed that at the extrema stability changes to instability or in other words masses upto 3 $M_\odot$ are indeed stable. Consequently stability requirements do not bring down the NS mass limits. A systematic study of ultracompact objects is under progress.

Note added in proof:

For further details see 'Ultracompact objects ($r < 3M$) in *General relativity*', B R Iyer, C V Vishveshwara and S V Dhurandhar 1983 (submitted for publication).

References

Bardeen J 1965 Ph.D. thesis, Caltech
Bardeen J M, Thorne K S and Meltzer D W 1966 *Ap. J.* **145** 505
Baym G, Bethe H and Pethick C 1971 *Nucl. Phy.* **A175** 225
Baym G, Pethick C and Sutherland P 1972 *Ap. J.* **170** 299
Buchdahl M 1959 *Phys. Rev.* **116** 1027
Chandrasekhar S 1931 *Astrophys. J.* **74** 81
Chandrasekhar S 1964 *Ap. J.* **140** 417
Chitre D M and Hartle J B 1976 *Ap. J.* **2** 597
Dhurandhar S V 1983 Talk in this symposium
Dhurandhar S V and Vishveshwara 1981 Unpublished

Hartle J B 1978 *Phys. Rep.* **46** 247
Kelley R L and Rappaport S 1980 in Pulsars (eds.) W Sieber and Wielebinski
Misner C W and Zapolsky H S 1964 *Phys. Rev. Lett* **12** 635
Oppenheimer J R and Volkoff G M 1939 *Phys. Rev.* **55** 374
Rhoades C and Ruffini R 1974 *Phys. Rev. Letts* **32** 324
Sabbadini A G and Hartle J B 1973 *Ap. and sp. sc.* **25** 117
Van Paradijs J 1979 *Ap. J.* **234** 609

Discussion

J. V. Narlikar: I don't see any justification for using the formula $L = 4\pi R^2 \sigma T^4$ for Black holes. It will be true for a flat space. But how can it be true for a highly curved space?

B. R. Iyer: I agree that there is no rigorous justification for using the above formula in the black hole context *e.g.* in deducing the time scale of black hole evaporation. However, for the van Paradij's result the arguments seem reasonable since one has taken the validity of the law in a local frame and put in all the relevant red shifts when referring to the arbitrary frame of the observer.

A. K. Raychaudhuri: You mention the condition that the Lagrangian change of pressure vanishes and explain it as the vanishing of p for a particular set of observers. But pressure is a scalar and if p vanishes, it should vanish for all observers. Is it not so?

B. R. Iyer: Yes. I agree. The comoving observers were only a convenient set of observers.

Neutrinos in gravitational collapse

S V DHURANDHAR
Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

1. Introduction

The work described in this paper has been done by Vishveshwara, Iyer and the present author (Dhurandhar and Vishveshwara 1981, 1983; Iyer *et al* 1983). In this work we concentrate on the general relativistic effects on neutrinos in the background geometry relevant to spherical gravitational collapse.

Recent considerations show that neutrinos have an important role to play in astrophysics. We shall be concerned here with the neutrinos emitted during a gravitational collapse of a star. In the relevant range of energies the geometric optics approximation is valid and the transport problem to a large extent can be investigated with the help of null geodesics. Since null geodesics form a major part of our considerations our analysis applies to other zero rest mass particles such as the graviton and the photon. In the low opacity limit our calculations are valid for the neutrinos which we assume are emitted from the interior of the object. Kembhavi and Vishveshwara (1980) examined the behaviour of neutrinos in static compact objects and showed that gravitational trapping of neutrinos occurs for an object whose radius is less than 1·5 times its Schwarzschild radius. In our case, we expect the trapping to be enhanced since the neutrinos will tend to be dragged inwards by the collapse.

The scenario is the following: A non-rotating spherical star undergoes collapse and emits zero rest-mass particles from its interior. We assume that the pressure is negligible as compared with the energy density so that the collapse is essentially geodetic. The interior metric in geometrised units ($c = 1, G = 1$) is described by the Friedmann dust line element,

$$\mathrm{d}S^2 = \mathrm{d}T^2 - S^2(T)\left[\frac{\mathrm{d}R^2}{1-\alpha R^2} + R^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\right], \tag{1}$$

where (R, θ, ϕ) are the comoving coordinates of each particle of the star, $S(T)$ the expansion factor which satisfies the differential equation,

$$(\mathrm{d}S/\mathrm{d}T)^2 = \alpha(1-S)/S \tag{2}$$

where $\alpha = 2m/R_b^3$, R_b is the R-coordinate of the particle on the boundary and m the geometrised mass of the object. By Birkoff's theorem the exterior geometry is Schwarzschild and is given by

$$\mathrm{d}S^2 = \left(1-\frac{2m}{r}\right)\mathrm{d}t^2 - \left(1-\frac{2m}{r}\right)^{-1}\mathrm{d}r^2 - r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2). \tag{3}$$

Since our geodesics emanate from a point inside the Friedmann geometry and emerge

into the Schwarzschild domain, our discussion may be broken up into three parts: (i) Null geodesics in the Friedmann geometry; (ii) the matching at the surface $R = R_b$ of the object, and (iii) Null geodesics in the Schwarzschild geometry.

2. Null trajectories

2.1 *Null geodesics in the interior geometry*

Since spherical symmetry is present in the problem, without loss of generality, we may choose the $\theta = \pi/2$ plane to study the behaviour of the geodesics. The first integrals can immediately be written down,

$$S\frac{dT}{d\lambda} = \Gamma, \quad R^2 S^2 \frac{d\phi}{d\lambda} = h; \quad ds = 0. \tag{4}$$

The constants Γ and h are the unscaled measures of the energy and angular momentum respectively of the particle and it is only the ratio $h/\Gamma = B$ (impact parameter) which determines completely a particular null geodesic. From (4) the radial propagation is given by

$$\frac{S^2}{1-\alpha R^2}(dR/dT)^2 = 1 - \frac{B^2}{R^2}. \tag{5}$$

The effective potential $1/R^2$ is repulsive so that all particles are forced to come out of the object.

It is convenient to treat the problem using the quantities χ and ψ defined as follows

$$S = \cos^2\chi, \; 0 \leqslant \chi \leqslant \pi/2 \text{ whence } T = \frac{1}{\sqrt{\alpha}}(\chi + \sin\chi\cos\chi).$$

The quantity ψ is the angle the tangent to the trajectory makes with the *outward* radial direction as measured by the comoving observer in the Friedmann geometry. From local differential geometry we have the relation, $B = R\sin\psi$ which is the same as the flat spacetime expression. Indeed, this is not surprising since the Friedmann geometry is conformal to the flat spacetime geometry and null geodesics are invariant under a conformal change in the metric. We now integrate (4) and (5). If the initial parameters are χ_0, ψ_0 and R_0 then the current coordinates are given by,

$$\chi = \begin{cases} \chi_0 + \frac{1}{4}(\chi_1(R_0) - \chi_1(R)) \text{ for } \psi_0 \leqslant \pi/2 \\ \chi_0 + \pi/4 - \frac{1}{4}(\chi_1(R_0) + \chi_1(R)) \text{ for } \psi_0 \geqslant \pi/2, \end{cases} \tag{6}$$

where,

$$\chi_1(R) = \sin^{-1}\left(\frac{1 + \alpha R_0^2\sin^2\psi_0 - 2\alpha R^2}{1 - \alpha R_0^2 \sin^2\psi}\right),$$

and

$$\phi = \begin{cases} \phi_0 - \frac{1}{2}(\phi_1(R) - \phi_1(R_0)) \text{ for } \psi_0 \leqslant \pi/2 \\ \phi_0 + \pi/2 + \frac{1}{2}(\phi_1(R) + \phi_1(R_0)) \text{ for } \psi_0 \geqslant \pi/2 \end{cases} \tag{7}$$

where,

$$\phi_1(R) = \sin^{-1}\left(\frac{1 + \alpha R_0^2 \sin^2\psi_0 - 2R_0^2/R^2 \sin^2\psi_0}{1 - \alpha R_0^2 \sin^2\psi_0}\right).$$

At $R = R_b$ we denote the values of χ and ϕ by χ_b and ϕ_b respectively.

2.2 *Matching conditions at the boundary $R = R_b$*

The angular coordinates θ and ϕ may be shown to be the same in both the geometries. This immediately provides us the relation,

$$r = RS(T). \tag{8}$$

Since our aim is to continue the null geodesic into the Schwarzschild domain the four partial derivatives $\partial r/\partial R$, $\partial r/\partial T$, $\partial t/\partial R$ and $\partial t/\partial T$ are needed to convert the components of the tangent vector to the null geodesic in the Friedmann geometry to the ones in the Schwarzschild geometry. The first two partial derivatives are directly obtained from (8) while the other two obtained by equating the line elements in the neighbourhood of the boundary. This procedure furnishes

$$\left.\frac{\partial t}{\partial T}\right|_{R=R_b} = \left(1 - \frac{2m}{r}\right)^{-1}(1 - \alpha R_b^2)^{1/2}$$

$$\left.\frac{\partial t}{\partial R}\right|_{R=R_b} = \frac{(1 - 2m/r)\,R_b S}{(1 - \alpha R_b^2)^{1/2}}\,(\mathrm{d}S/\mathrm{d}T). \tag{9}$$

The Schwarzschild time t for the particle on the boundary can be given unambiguously if we assume that the collapse begins at $t = 0$, that is, $t = 0$ at $T = 0$, then,

$$t = \left(\frac{R_b}{2m} - 1\right)^{1/2}(R_b + 4m)\chi + R_b\left(\frac{R_b}{2m} - 1\right)^{1/2}\sin\chi\cos\chi$$

$$+\, 2m\ln\left(\frac{(R_b/2m - 1)^{1/2} + \tan\chi}{(R_b/2m - 1)^{1/2} - \tan\chi}\right) \tag{10}$$

2.3 *Null geodesics in the Schwarzschild region*

This part of the discussion has been amply dealt with in the literature before but for the sake of completeness we only mention the results.

The Killing symmetries of the spacetime provide the following two first integrals

$$\left(1 - \frac{2m}{r}\right)\frac{\mathrm{d}t}{\mathrm{d}\lambda} = \bar{l} \qquad r^2\frac{\mathrm{d}\phi}{\mathrm{d}\lambda} = \bar{h}. \tag{11}$$

Since the geodesic is null the third integral is $\mathrm{d}s = 0$. The radial equation is then given by,

$$\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 = h^2\left(1 - \frac{2m}{r}\right)\left[\frac{1}{b^2} - l^2(r)\right], \tag{12}$$

where $b = \overline{h}/\overline{\Gamma}$ and

$$V^2 = \frac{1}{r^2}\left(1 - \frac{2m}{r}\right)$$

is the effective potential. The effective potential possesses a maximum of height $1/27m^2$ at $r = 3m$ and decays as $r \to \infty$. Therefore, the particles having $b > 3\sqrt{3}m$ are confined to the side of the potential barrier in which they are initially present. However, when $b < 3\sqrt{3}m$ the potential does not present any barrier.

Since the null geodesics are dependent solely on the impact parameters b and B, it is necessary to obtain b in terms of the initial parameters in the interior geometry. The relation may be deduced from the equation,

$$\overline{k}^0 = \frac{\partial t}{\partial T} k^0 + \frac{\partial t}{\partial R} k^1, \tag{13}$$

where $\overline{k}^\alpha$, $\alpha = 0, 1, 2, 3$ is the tangent vector to the null geodesic in the interior geometry while $\overline{k}^\alpha$, $\alpha = 0, 1, 2, 3$ is the corresponding tangent vector in the Schwarzschild domain. Equation (13) immediately furnishes the required relation,

$$b = \frac{R_0 \sin\psi_0 \cos^2\chi_b}{[(1 - \alpha R_b^2)^{1/2} - \sqrt{\alpha}\tan\chi_b (R_b^2 - R_0^2 \sin^2\psi_0)^{1/2}]}. \tag{14}$$

3. Spectral shift

The spectral shift plays a pivotal role in the energy considerations of the flux. If ν_e is the energy of the emitted particle as measured by a comoving observer in the Friedmann geometry and ν_{ob} the observed energy as measured by the static observer in the Schwarzschild geometry then the spectral shift is given by the formula,

$$1 + z = \nu_e/\nu_{ob}. \tag{15}$$

In terms of the components of the tangent vectors k^α and $\overline{k}^\alpha$, $\nu_e = k^0$ while

$$\nu_{ob} = \left(1 - \frac{2m}{r}\right)^{1/2} \overline{k}^0.$$

In view of (13) the spectral shift is given by,

$$1 + z = (1 + z_c)(1 + z_d)(1 + z_s), \tag{16}$$

where

$$1 + z_c = (\cos^2\chi_b/\cos^2\chi_0),$$

$$1 + z_d = [(1 - \alpha R_b^2)^{1/2} - \alpha^{1/2}\tan\chi_b (R_b^2 - R_0^2\sin^2\psi_0)^{1/2}]^{-1},$$

$$1 + z_s = \left(1 - \frac{2m}{r}\right)^{1/2}.$$

It is seen that the spectral shift $1 + z$ can be thought to be made up of three portions $1 + z_c$, $1 + z_d$ and $1 + z_s$. The quantity $1 + z_c$ may be referred to as the cosmological blue shift. It is known that in an expanding universe particles lose energy; the reverse is to be

expected in the collapsing situation. The spectral shift $1+z_d$ is the Doppler shift due to the relative velocity between the comoving observer in the Friedmann geometry on the surface of the star and the static observer in the Schwarzschild domain instantaneously coinciding with him. The portion $1+z_s$ is the Schwarzschild gravitational red-shift which is due to the particle losing its energy in climbing out of the gravitational potential well.

4. Backward emission

The notion of backward emission is not new in the collapsing situation. Jaffe (1969) had discussed this aspect in connection with the optical appearance of spherically symmetric collapsing star. Let us consider a particle to be emitted from the surface of the collapsing star. The concept can be extended in an obvious way to an interior point in the star. Consider the particle to be emitted with $\psi_0 = 0$ (emitted radially outward). This particle will have $dr/d\lambda|_{\text{boundary}} > 0$. However, the particle emitted with $\psi_0 = \pi/2$, emitted tangentially to the surface as seen by the comoving observer will have

$$dR/d\lambda|_{R=R_b} = 0.$$

But since the surface is collapsing the particle possesses a *negative* radial component of velocity as seen by the static observer in the Schwarzschild geometry and therefore $dr/d\lambda|_{\text{boundary}} < 0$. Such a particle having $dr/d\lambda|_{\text{boundary}} < 0$ is said to be backward emitted. Otherwise the particle is forward emitted. Since $dr/d\lambda$ is a continuous function of the emission angle ψ_0, there is a value of $\psi_0 = \psi_m = \pi/2$ for which $dr/d\lambda|_{\text{boundary}} = 0$. For $\psi_0 < \psi_m$ the neutrinos are forward-emitted. Now the particle with $\psi_0 = \pi$ is also forward emitted (if the blackhole has not been formed) and hence again applying continuity arguments there is a value of $\psi_0 = \psi'_m > \pi/2$ for which $dr/d\lambda|_{\text{boundary}} = 0$. Recapitulating, the following is the situation. For $\psi_m < \psi_0 < \psi'_m$ the particles are backward emitted while for $0 < \psi_0 < \psi_m$ and $\psi'_m < \psi_0 < \pi$ the particles are emitted in the forward direction. In the three dimensional picture ψ_m and ψ'_m are the half angles of the cones which demarcate the directions of forward emission from those of backward emission.

The arguments given above can be extended to the interior easily but care must be taken to check whether backward emission occurs at all. For a point on the surface backward emission must occur and hence by continuity, backward emission must occur for points near the surface. However, if emission occurs from the centre the particles in all directions travel radially and hence are all forward emitted. Therefore there is a value of the radius $R = R_1$ such that for $R_0 < R_1$ there is only forward emission. This behaviour can be conveniently studied with the help of the function F of the initial parameters

$$F = \left(1 - \frac{2m}{R_b}\right)\left(1 - \frac{R_0^2}{R_b^2}\sin^2\psi_0\right) - \frac{2m}{R_b}\tan^2\chi_b \tag{17}$$

$F > 0$ implies that the particle is forward emitted while $F < 0$ signifies backward emission. The function is nothing but the $dr/d\lambda|_{\text{boundary}}$ multiplied by a positive factor to remove square roots. We have the following properties of F.

(i) $\psi_0 = 0$ or π we have $F > 0$ if $r_b > 2m$. Radially directed neutrinos are forward-emitted.

(ii) For $R_0 = 0$, F is a positive. All neutrinos emitted from the centre of the star are forward-emitted.

(iii) For $R = R_b$, $\psi_0 = \pi/2$ we have $F < \dot{0}$ which means that a particle emitted tangentially to the surface is backward emitted.

5. The escape and confinement of the zero mass particles from infinity.

If the particle at some time (however large) can be detected at a point whose r-coordinate is arbitrarily large, then such a particle is said to have escaped to infinity; otherwise we say that it is confined. In the confined state the particle may remain bound in the vicinity of the object or fall into the blackhole. The confinement or the escape of the particle is governed by three functions, namely, $r_b = R_b \cos^2 \chi_b$, F and b all of which are functions of the three initial parameters χ_0, R_0 and ψ_0. The truth table describes the confinement process in a lucid way. The confinement or the escape of the particle is determined from the truth and falsity of the following three statements (A) $r_b < 3m$; (B) $F < 0$; (C) $b < 3\sqrt{3}m$.

Let D be the condition that the neutrino is confined from infinity. The symbol 1 denotes the truth of the statement and 0 the falsity in table 1. There are eight possibilities and as can be seen from the table the particle is confined in four possible ways. If one actually assigns these values to the conditions A, B, C and D, then D can be written as an arithmetic expression in A, B and C.

$$D = A(1-C) + BC. \tag{18}$$

For example the fourth possibility $A = 1$ and $B = C = 0$ means that $r_b < 3m$, $F > 0$ and $b > 3\sqrt{3}m$. In this case the particle is forward-emitted, hits the effective potential barrier from the 'inside' and falls back into the eventually formed blackhole.

The conditions A, B and C divide the (R_0, χ_0, ψ_0) space into regions in which D is either true or false. Clearly R_0, χ_0 and ψ_0 determine a unique null geodesic. But a point R_1, χ_1 and ψ_1 lying on the null geodesic can equally well be used to describe the same null geodesic. The null geodesics therefore furnish an equivalence relation on the space of initial parameters. In particular we may choose the initial parameters when the null

Table 1. Truth table for neutrino confinement

$r_b < 3m$ (A)	$F < 0$ (B)	$b < 3\sqrt{3}m$ (C)	Confinement (D)
1	1	1	1
1	1	0	1
1	0	1	0
1	0	0	1
0	1	1	1
0	1	0	0
0	0	1	0
0	0	0	0

geodesic reaches the boundary $R = R_b$ and then relate the results to an interior point $R_0 = R_b$.

5.1 *The surface case* $R_0 = R_b$

Since the functions r_b, F and b determine the confinement of the particles it seems necessary to adopt a scheme which displays these functions simultaneously and then study their relative behaviour. We have already fixed $R_0 = R_b$, now we fix χ_b for the time being so that the only variable is ψ_0. W note that r_b is a monotonic function of ψ_0 and hence b may be treated as a function of r_b. The function $b(r_b)$ defines a curve in the (r_b, b) plane henceforth referred to as the b-curve. The condition $F = 0$ determines an equation in r_b which may be solved. The $F = 0$ condition can be displayed by vertical lines which demarcate the forward emitted $(F > 0)$ particles from the backward emitted $(F < 0)$ ones. We now possess all the machinery to describe the confinement process at various stages of the collapse, as the collapse progresses. During the early stages there is no confinement since $b \sim r_b$ and $r_b > 3m$. For confinement to occur it is necessary that r_b be less than $3m$. With this information one can obtain a rough upper bound on the radius of the object when the particle was emitted for confinement to occur. The upper bound is simply obtained by setting $R_b \cos^2 \chi_b = 3m$, $\psi_0 = \pi$ and then solving for χ_0. The upper bound turns out to be about $9{\cdot}9m$ for $R_b/m \gg 1$. For a value of χ_0 little greater than that corresponding to the upper bound, the b-curve enters the $r_b < 3m$ region but the curve still remains below the $b = 3\sqrt{3}m$ line. Since F is positive there is no confinement. Confinement occurs if the epoch is advanced a little more that b becomes greater than $3\sqrt{3}m$ and the b-curve intersects the $b = 3\sqrt{3}m$ line at two points. Within these two values of r_b, there is an interval on the r_b-axis which corresponds to an interval on the ψ_0 axis for which the particles are confined. In the three-dimensional picture the confinement process starts with the confinement occurring between two cones. For more advanced epochs the interval of confinement grows larger in both directions along the ψ_0-axis with several of the confinement conditions coming into play. The cores encompass a larger range of directions with the outer cone opening out and the inner cone closing in. When the b-curve touches the $r_b = 2m$ line the inner cone degenerates into a line $\psi_0 = \pi$ and only a single cone remains. Till this instant we call the confinement as the double cone confinement. After this epoch further advancement makes the cone of directions grow larger until total confinement in all directions occurs when the size of the object reaches $2m$.

The entire process of confinement as a function of the epoch of emission can be shown in the (ψ, χ) plane as is done in figure 1. A neutrino emitted with the parameters $R_0 = R_b$, $\chi = \chi_0$ and $\psi = \psi_0$ will not escape to infinity if the point lies in the region above the undotted curve S. The presence of the minimum to the curve implies the existence of double cone confinement. This may be seen as follows: Any given epoch $\chi = \chi_0$ represents a horizontal line in the (ψ, χ) plane. Confinement in some directions will occur at this epoch only if the line $\chi = \chi_0$ crosses the curve S. In the early stages of the collapse this line is below the curve S. As the collapse progresses this line moves upward until it touches the curve S at its minimum and the confinement process begins. A little later the line intersects the curve S at two points ψ_1 and ψ_2. In the directions $\psi_1 < \psi < \psi_2$ (double cone) the neutrinos are confined from infinity. At a still later stage the line intersects the curve S at only one point and only a single cone of confinement directions exists.

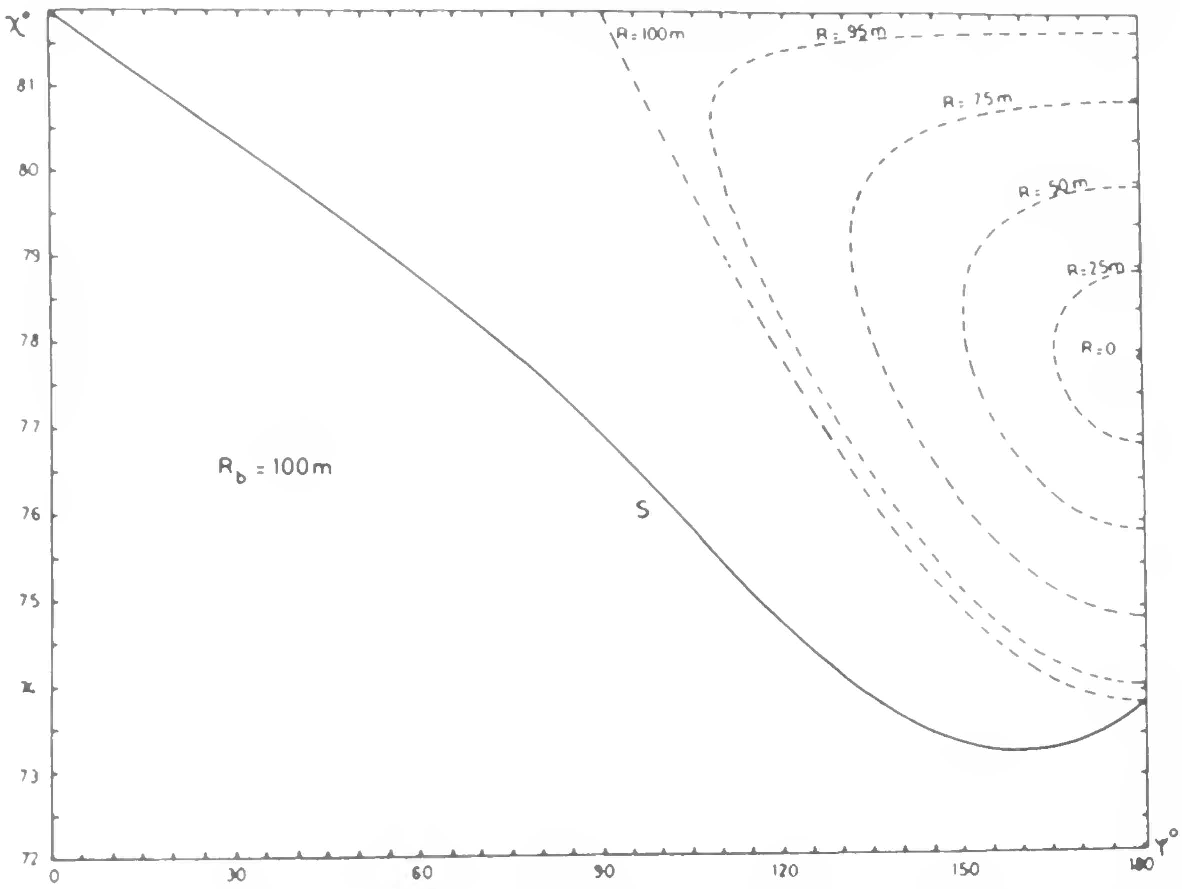

Figure 1. The confinement process is depicted for a wide range of the initial parameters (R_0, ψ_0, χ_0) and $R_b = 100$ m. The unbroken curve S corresponds to the confinement process of neutrinos emitted from the surface $R_0 = R_b$. For the choice of the parameters χ_0, ψ_0 and $R_0 = R_b$ the neutrino is confined if the point (ψ_0, χ_0) lies above the curve S. The dashed curves describe the confinement process for the interior points $R_0 < R_b$. The curves are drawn for the fixed value of $\chi_0 = 81{\cdot}87°$ obtained from the equation $R_b \cos^2 \chi_0 = 2m$.

5.2 *The general case of* $R_0 < R_b$

To examine the confinement of the particles emitted from an interior point it is only necessary to relate the initial parameters in the interior to those on the boundary. Two choices exist in this situation since the null geodesic cuts the boundary at two points, the past and the future. If χ_{R_0} and ψ_{R_0} are the values of χ and ψ respectively, where the null geodesic cuts the boundary in the past then the particle is confined if χ_{R_0} and ψ_{R_0} lie in the region above the undotted curve S. The parameters χ_{R_0} and ψ_{R_0} are related to the actual parameters by the relations,

For $\quad \psi_0 \leqslant \pi/2$,

$$\chi_{R_0} = \chi_0 - \pi/4 + \tfrac{1}{4}[\chi_1(R_0) - \chi_1(R_b)].$$

For $\quad \psi_0 \geqslant \pi/2$,

$$\chi_{R_0} = \chi_0 - \tfrac{1}{4}[\chi_1(R_0) - \chi_1(R_b)],$$

where

$$\chi_1(R) = \sin^{-1}\left(\frac{1 + \alpha R_0^2 \sin^2 \psi_0 - 2\alpha R^2}{1 - \alpha R_b^2 \sin^2 \psi_0}\right),$$

and

$$\psi_{R_0} = \sin^{-1}\left(\frac{R_o}{R_b} \sin \psi_0\right) \text{ with } \psi_{R_0} \geqslant \pi/2. \tag{19}$$

For a fixed value of R_0 and χ_0 the equations give a parametric representation of a curve. The dotted curves in figure 1 are drawn for various values of R_0 with χ_0 chosen arbitrarily ($R_b \cos^2 \chi_0 = 2m$). The exact value of χ_0 is not very important to discuss the confinement process because χ_0 is merely an additive constant in (19). Any change in χ_0 merely displaces the dotted curve in the vertical direction.

The curves are *convex* and hence the confinement process for an interior point also begins with a double cone which gives way to a single cone at a later stage. For a fixed value of χ_0 and for various values of R_0 the curves are nested with the inner curves corresponding to smaller values of R_0. This shows that a confinement process begins later and total confinement occurs earlier for a point lying more to its interior.

This terminates the discussion of the null trajectories.

6. Energy flux profiles observed at infinity

We apply the earlier considerations to compute the energy flux as received by an observer sufficiently removed from the collapsing star. We make the following assumptions about the emission of the particles. (i) Emission of the particles takes place in a 'flash' and uniformly over a spherical shell. (ii) Each source on the shell emits isotropically as seen by the comoving observer in the Friedmann geometry.

In general the arrival time t is a function of the three initial parameters but our assumptions imply that R_0 and χ_0 are fixed so that t becomes a function of ψ_0 only. The flux profile then depends crucially on the function $t(\psi)$. The flux profile $I(t)$ can be obtained by geometric considerations. Figure 2a shows the trajectories in the (r, ϕ) plane with emission taking place on the shell $R_0 = R_b$. If $R_0 < R_b$ the geometry will be somewhat more complicated but the derivation will essentially remain unaffected. Consider a small elementary area AB of height $\delta\phi$ at the observer's position r_0 and consider all the particles passing through AB between the time instants t and $t + \Delta t$. The null geodesics which reach A at t and B at $t + \Delta t$ mark off a ring $A'B'$ on the shell whose

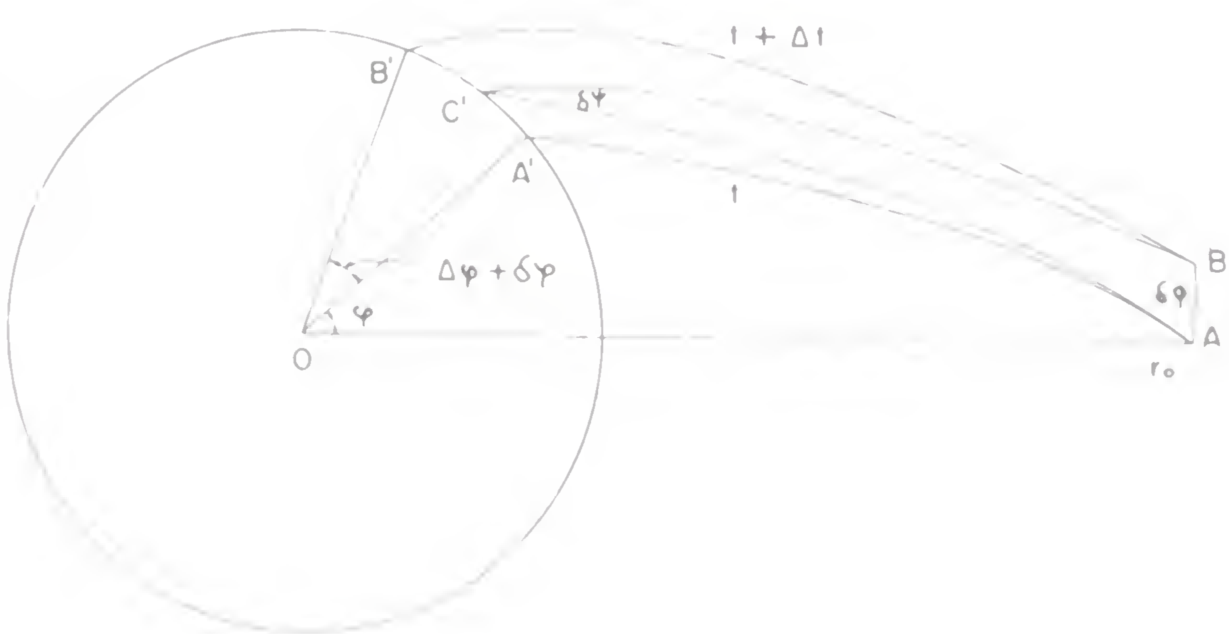

Figure 2(a). Null geodesics emanating from the points A' and B' lying on the shell $R = R_b$ and which reach A and B at times t and $t + \Delta t$ respectively. AB is the elemental area at $r = r_0$ at the point of observation. C' is a typical source which contributes to the flux at AB between the times t and $t + \Delta t$. The angles $\Delta\phi$ and $\delta\phi$ subtended by the arc $A'B'$ at O and $\delta\psi$ have been exaggerated for the sake of clarity in the diagram.

width is $\Delta\phi$. C' is a typical point on the ring which contributes to the flux. Then by working out the geometry the flux $I(t)$ can be written as,

$$I(t)\Delta t \propto \frac{2\pi \sin\psi\,\delta\psi}{2\pi \sin\phi\,\delta\phi}\cdot\frac{2\pi \sin\phi\,\Delta\phi}{1+z}. \tag{20}$$

Taking limits as both Δt and $\delta\phi$ tend to zero we get

$$I(t) = \frac{\sin\psi}{(1+z)\mathrm{d}t/\mathrm{d}\psi}, \tag{21}$$

where we have assumed the proportionality constant to be unity. The above formula was obtained on the basis of t being a monotonic function of ψ. But this is not in general the case especially for advanced epochs and the formula then needs to be modified. Fortunately, the modification is simple, since one value of t corresponds to at most two values of ψ. In this case the total flux is just the sum of the fluxes corresponding to ψ_1 and ψ_2,

$$I(t) = \left[\frac{\sin\psi}{(1+z)|\mathrm{d}t/\mathrm{d}\psi|}\right]_{\psi=\psi_1} + \left[\frac{\sin\psi}{(1+z)|\mathrm{d}t/\mathrm{d}\psi|}\right]_{\psi=\psi_2}. \tag{22}$$

There are essentially four types of behaviour of the flux profile $I(t)$ depending on the epoch at which the emission occurs.

6.1 *The early stage*

The early stage can be defined by the condition that

$$\left.\frac{\mathrm{d}^2 t}{\mathrm{d}\psi^2}\right|_{\psi=\pi} < 0$$

and holds for early epochs. Figure 2b shows both the functions $t(\psi)$ and $I(t)$ which are monotonically increasing functions of their respective arguments. In this case there is no confinement or any delay in the arrival time due to the particles circling around the collapsing object. The gravitational effects are minimal in this case. There are no bursts or decays in the flux profile.

6.2 *The burst*

For more advanced epochs $\mathrm{d}^2 t/\mathrm{d}\psi^2|_{\psi=\pi}$ changes sign and becomes positive. A typical curve $t(\psi)$ is shown in figure 2(c). The function $t(\psi)$ attains a maximum at $t = t_m$ and remains finite for the entire range of ψ. The flux profile possesses two distinctive features. (i) A burst which occurs at the end of the flux profile. (ii) A discontinuous increase in the flux at $t = t_\pi$ where t_π is the arrival time of the particle emitted with $\psi = \pi$.

The reason for the burst is that $\mathrm{d}t/\mathrm{d}\psi$ vanishes at the maximum of the curve at $t = t(\psi)$. It may be remarked that the burst is the reflection of the fact that we have chosen the emission to occur in a flash. The discontinuity in the flux function occurs due to the following reason: For $t < t_\pi$ the particles arriving at r_0 are emitted from a single ring on the shell. However, when $t > t_\pi$ there are two values of ψ which correspond to the same value of t and hence two corresponding values of ϕ. This means that there are

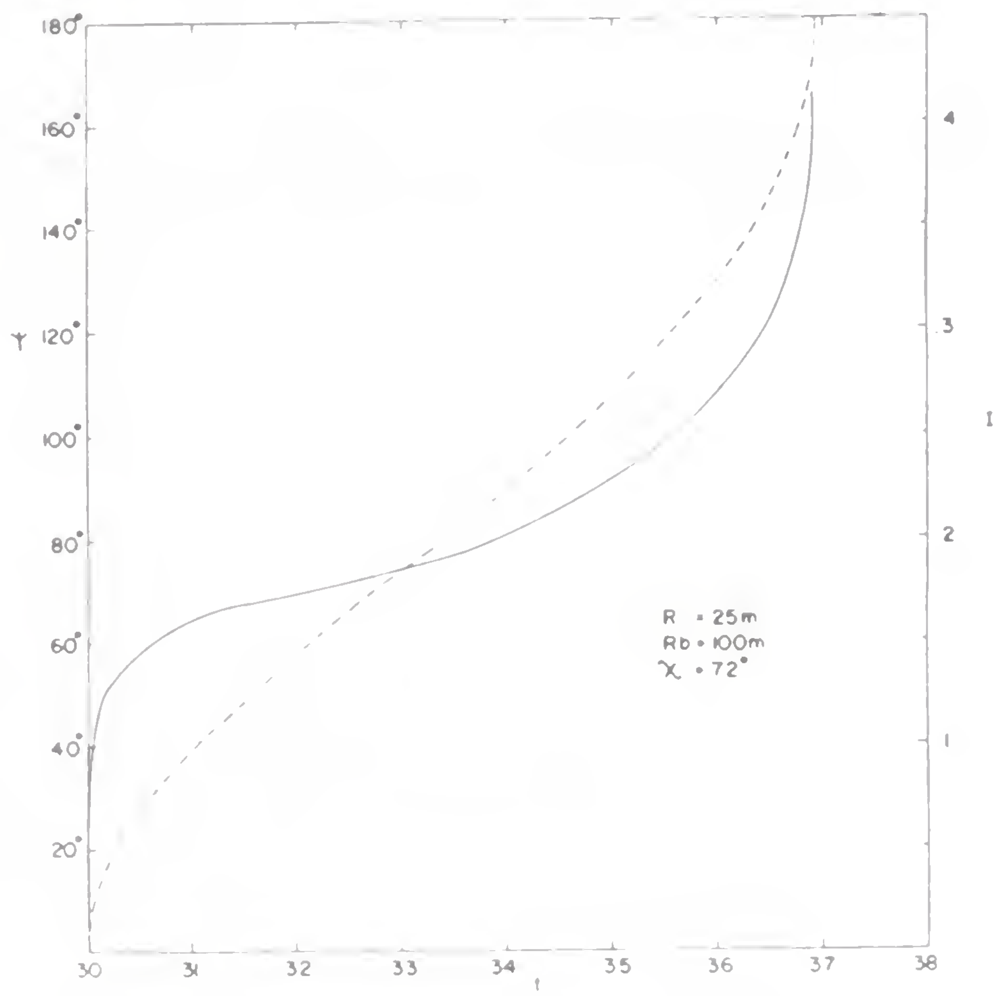

Figure 2(b). The undashed curve $I(t)$ is depicted for the early stages of the collapse. The flux intensity is seen to be a monotonic function of the arrival time t. The dashed curve shows the ψ–t correlation.

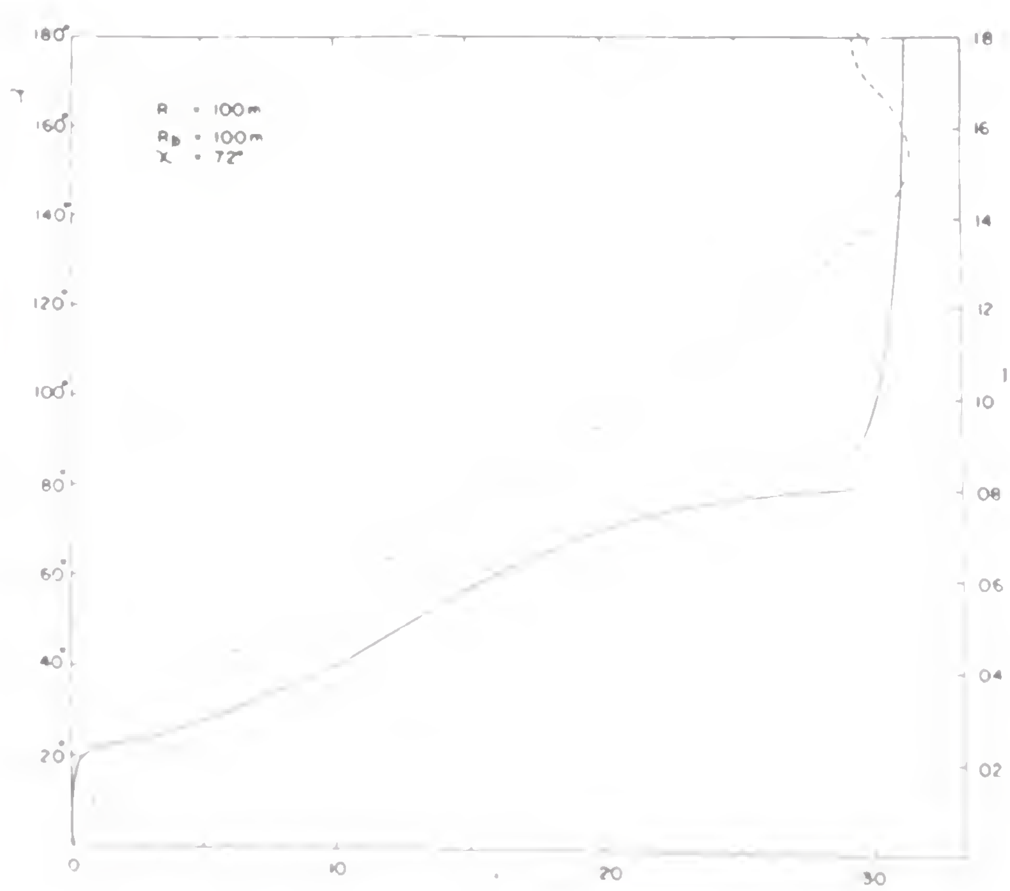

Figure 2(c). Dashed curve $t(\psi)$ develops a maximum but remains finite. The undashed curve $I(t)$ has a discontinuity and terminates in a burst.

two rings on the shell which contribute to the total flux. The sudden contribution of flux of the second ring at $t = t_\pi$ causes the discontinuity.

6.3 *The double cone*

Let χ_{dc} (the subscript for double cone) be the epoch at which the confinement process begins. We recall that the confinement process begins with the confinement directions being enclosed between two cones. Lt χ_{sc} (the subscript for single cone) be the epoch when the inner cone degenerates into the line $\psi = \pi$. In this case we consider the flux profiles for χ satisfying $\chi_{dc} < \chi < \chi_{sc}$. The curves $t(\psi)$ and $I(t)$ are depicted in figure 2(d). We observe that the function $t(\psi)$ has two asymptotes at $\psi = \psi_1$, and $\psi = \psi_2$, which are the half angles of the two cones. The time of arrival tends to infinity as ψ approaches either ψ_1 and ψ_2, due to the particles circling around the object at $r = 3m$.

One observes that the burst is absent. However, as the particles emitted with $\psi = \pi$ still reach the observer, the discontinuity is present but is delayed since t_π is an increasing function of the epoch. In fact as χ approaches χ_{sc}, t_π tends to infinity and the discontinuity is delayed without limit.

It may be remarked that this phase of collapse remains for a very small fraction of time taken for the entire collapse. For $R_b \sim 100\, m$, $\chi_{sc} - \chi_{dc} \sim 0{\cdot}01$ radians.

6.4 *The single cone*

The case deals with the epoch $\chi > \chi_{sc}$ when only a single cone exists and involves the last stages of the collapse. The arrival time t is a monotonic increasing function of ψ and tends to infinity as ψ approaches ψ_1. For $\psi > \psi_1$, the particles are swallowed up by the incipient black hole. Figure 2(e) shows the function $t(\psi)$ together with flux profile $I(t)$. Initially the flux increases in value until it reaches a maximum and then gradually

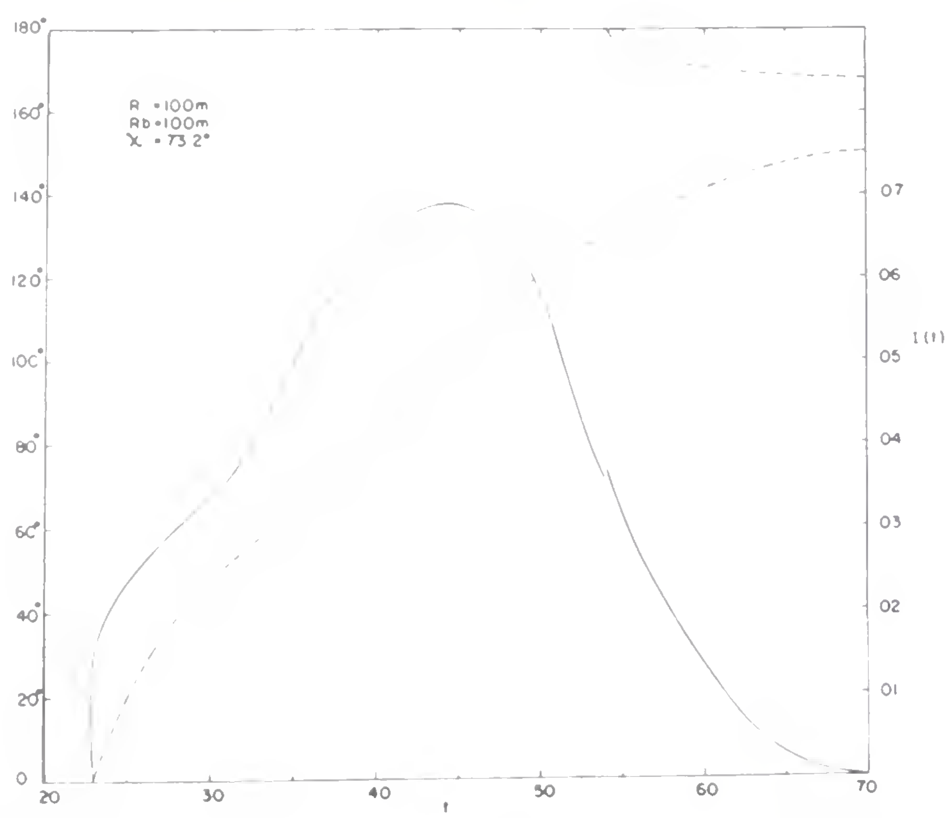

Figure 2(d). The dashed curve $t(\psi)$ tends to infinity as ψ approaches the critical values ψ_1 and ψ_2. The undashed curve $I(t)$ possesses a maximum and a discontinuity. The flux strength decays as t tends to infinity.

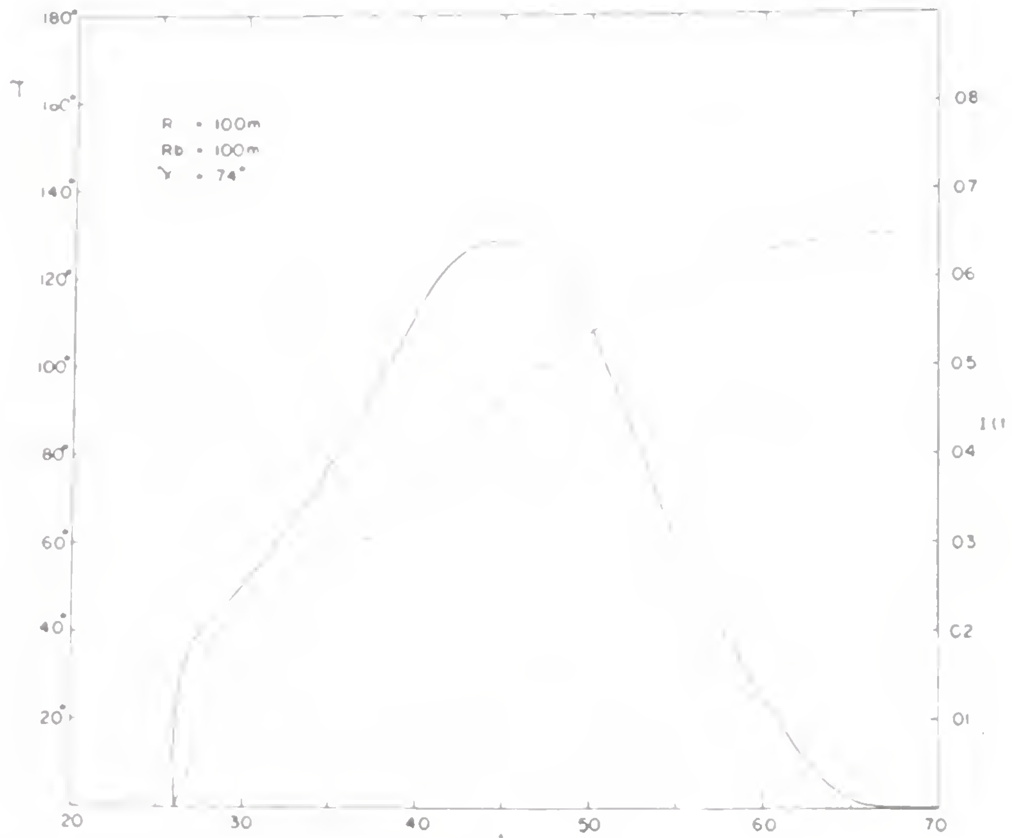

Figure 2(e). The dashed curve $t(\psi)$ tends to infinity as ψ approaches the critical value ψ_1. The flux intensity develops a maximum and then decays for large times t.

decays for large values of t. Both the features of discontinuity and the burst are absent in this case.

7. The Dirac formalism

Till now the treatment has been classical in that the calculations involved the null geodesics. The problem can be treated at a little deeper level considering zero mass perturbations on the background geometry. We restrict ourselves to the neutrino case and set up the necessary Dirac formalism for the massless particle. Our program will be to first write the Dirac equation in both the interior and the exterior geometry and then obtain solutions. Finally the solutions will be matched at the boundary $R = R_b$ so that a complete solution is obtained in the entire spacetime.

The Dirac equation in curved spacetime may be written as

$$\gamma^a \nabla_a \psi = 0 \tag{23}$$

where γ^a are the flat spacetime 4×4 Dirac matrices satisfying

$$\gamma^a \gamma^b + \gamma^b \gamma^a = 2\eta^{ab}, \tag{24}$$

η^{ab} being the flat spacetime metric tensor and,

$$\nabla_a \psi = e^\mu_a(\partial_\mu - \Gamma_\mu)\psi, \tag{25}$$

where e^μ_a are the chosen tetrad fields satisfying the conditions

$$\eta^{ab} e^\mu_a e^\nu_b = g^{\mu\nu}. \tag{26}$$

The Γ_μ are the spinor affine connections and can be given by an explicit formula,

$$\Gamma_\mu = -\tfrac{1}{4}\gamma^a \gamma^b e^\nu_a e_{b\nu;\mu} \tag{27}$$

The spin of the neutrino is polarised antiparallel to its momentum and the neutrino satisfies an additional helicity condition, namely,

$$(1+i\gamma_5)\psi = 0, \tag{28}$$

where

$$\gamma_5 = \frac{\varepsilon^{\mu\nu\alpha\beta}\gamma_\mu\gamma_\nu\gamma_\alpha\gamma_\beta}{4!\sqrt{-g}}, \quad \gamma_\mu = e_{a\mu}\gamma^a.$$

7.1 *Solutions in the interior geometry*

The choice of the tetrad e^μ_a with the non-vanishing components,

$$e^T_0 = 1, \qquad e^\phi_2 = (RS\sin\theta)^{-1},$$
$$e^\theta_1 = (RS)^{-1}, \quad e^R_3 = R^{-1}(1-\alpha R^2)^{-1/2},$$

enables us to write the Dirac equation on the interior geometry,

$$\left[RS\gamma^0\left(\partial_T + \frac{3}{2}\frac{1}{S}\frac{\mathrm{d}S}{\mathrm{d}T}\right) + R(1-\alpha R^2)^{1/2}\gamma^3\left(\partial_R + \frac{1}{R}\right)\right.$$
$$\left. + \gamma^1(\partial_\theta + \tfrac{1}{2}\cot\theta) + \gamma^2 \operatorname{cosec}\theta\,\partial_\phi\right]\psi = 0. \tag{29}$$

These are the four coupled equations in the four components of ψ and can be solved using the method of separation of variables. After a fair amount of computation we obtain the solutions in the WKB approximation ($k \gg 1$, $\omega \gg 1/\sqrt{\alpha}$) as,

$$\psi^{i(1)} = u^{i(1)}\left(\exp i\omega\int_{R_b}^{R}\frac{q\,\mathrm{d}R}{(1-\alpha R^2)^{1/2}}\right),$$
$$\psi^{i(2)} = u^{i(2)}\exp\left(-i\omega\int_{R_b}^{R}\frac{q\,\mathrm{d}R}{(1-\alpha R^2)^{1/2}}\right), \tag{30}$$

where $q(R) = \left(1 - \dfrac{k^2}{\omega^2 R^2}\right)^{1/2}$,

$$u^{i(1)} = \frac{\exp -i(\omega\tau - m\phi)}{2RS^{3/2}}(a_+S_1, a_-S_2, a_+S_1, a_-S_2)^T,$$

$$u^{i(2)} = \frac{\exp -i(\omega\tau - m\phi)}{2RS^{3/2}}(a_-^*S_1, a_+^*S_2, a_-^*S_1, a_+^*S_2)^T,$$

and

$$a_\pm(q) = \frac{1}{\sqrt{q\omega}}\left(\frac{k}{\omega R} + i(q\pm 1)\right).$$

The angular functions S_1 and S_2 are spin-weighted spherical harmonics of spin weight half. The tetrad has been chosen for the comoving observer and this spinor ψ^i is tied to this tetrad. The helicity condition has been used and it forces ψ to be of the form $(\eta, \eta)^T$ where η itself is a two-component object. The quantity τ is nothing but 2χ. Since ϕ is Killing coordinate the dependence of the spinor on ϕ is of a simple nature. The

superscripts 1 and 2 indicate the outgoing and the incoming nature of the solutions respectively. The superscript i denotes that the solution pertains to the interior metric.

7.2 *The solutions in the exterior geometry*

The procedure followed to obtain solutions in the Schwarzschild metric is similar to that of the interior case. The tetrad corresponds to that of the static observer. Its nonvanishing components are given by,

$$e^{t}_{0'} = \left(1 - \frac{2m}{r}\right)^{-1/2}, \quad e^{\phi}_{2'} = (r \sin\theta)^{-1/2},$$

$$e^{\theta}_{1'} = 1/r, \qquad e^{r}_{3'} = \left(1 - \frac{2m}{r}\right)^{1/2},$$

where the prime and the bars represent quantities pertaining to the exterior geometry. The Dirac equation is

$$\left[\left(1-\frac{2m}{r}\right)^{-1/2}\gamma^0 \partial_t + \left(1-\frac{2m}{r}\right)^{1/2}\left(\partial_r + \left(1-\frac{2M}{r}\right)^{-1/2}\frac{M}{2r^2} + \frac{1}{r}\right)\gamma^3\right.$$
$$\left. + \frac{1}{r}(\gamma^1(\partial_\theta + \tfrac{1}{2}\cot\theta) + \gamma^2 \operatorname{cosec}\theta\, \partial_\phi)\right]\psi = 0. \tag{31}$$

The solutions as in the previous case are given by

$$\psi^{e(1)} = u^{e(1)}(\bar{q}) \exp i\bar{\omega} \int_{r_b^*}^{r^*} \bar{q}\, dr^*,$$

$$\psi^{e(2)} = u^{e(2)}(\bar{q}) \exp - i\bar{\omega} \int_{r_b^*}^{r^*} \bar{q}\, dr^*, \tag{32}$$

where,

$$u^{e(1)}(\bar{q}) = \frac{\exp - i(\bar{\omega}t - \bar{m}\phi)}{2r\left(1-\frac{2m}{r}\right)^{1/4}}(\bar{a}_+ S_1, \bar{a}_- S_1, \bar{a}_+ S_2, \bar{a}_- S_2)^T,$$

$$u^{e(2)}(\bar{q}) = \frac{\exp - i(\bar{\omega}t - \bar{m}\phi)}{2r\left(1-\frac{2m}{r}\right)^{1/4}}(\bar{a}^*_- S_1, \bar{a}^*_+ S_2, \bar{a}^*_- S_1, \bar{a}^*_+ S_2)^T,$$

$$\bar{q} = \left[1 - \frac{\bar{k}^2}{\bar{\omega}^2 r^2}\left(1-\frac{2m}{r}\right)\right]^{1/2}$$

$$\bar{a}_\pm(q) = \frac{1}{(|\bar{q}\omega|)^{1/2}}\left[\frac{\bar{k}\left(1-\frac{2m}{r}\right)^{1/2}}{\bar{\omega}r} + i(\bar{q} \pm 1)\right].$$

The superscript e corresponds to the exterior solution. The coordinate r^* is the well-known tortoise coordinate.

Now it remains only to match these two solutions at the boundary.

7.3 *The junction conditions*

We assume that we are given the solutions in the interior geometry. Our aim will be to continue these solutions to the exterior geometry. We observe that the interior solution ψ^i is tied up to the tetrad along the (T, R) directions whereas the exterior solution ψ^e is tied to the tetrad along the (t, r) coordinates. Therefore to accomplish the matching, it is necessary to appropriately rotate the ψ^i by a matrix $\mathscr{s}$ corresponding to the Lorentz transformation between the two tetrads.

$$\psi^e = \mathscr{s}\psi^i. \tag{33}$$

In accordance with the theory of partial differential equations, it is not enough to match only the values of ψ at the boundary but it is also necessary to match the derivatives. The derivatives are matched in the following manner,

$$\nabla_{\bar{v}}\psi^e = \frac{\partial x^\mu}{\partial \bar{x}^v}\mathscr{s}\nabla_\mu \psi^i. \tag{34}$$

Since $\mathscr{s}$ appears in this equation also, the tangent space to ψ^i are not merely continued across the boundary but there is a 'mixture' of these tangent spaces that is continuous. In fact (34) defines a one-one mapping of the tangent spaces and hence determines a unique solution in the exterior geometry once the interior solution is given. This jumbling of the components at the boundary occurs because the Dirac equations in the two regions are not connected merely by a coordinate transformation owing to the choice of the representations of γ^μ's. The angular parts of the solutions immediately yield, $\bar{k} = k$ and $\bar{m} = m$. The matching of the rest of the solution gives two equations to determine $\bar{\omega}$. The two equations arise from the matching of the two components of ψ. In order that a unique $\bar{\omega}$ may be obtained the equations must be consistent. The relation for $\bar{\omega}$ in terms of the quantities belonging to the interior region is,

$$\bar{\omega} = \frac{\omega}{S_b}((1-\alpha R_b^2)^{1/2} - \sqrt{\alpha}R_b \tan\chi_b q). \tag{35}$$

The consistency requirement gives rise to the backward emission effect in a most natural and elegant way. It is possible to match the outgoing interior solution with the outgoing exterior solution only for $B < B_m$ where $B = k/\omega$ and B_m is the root of the equation,

$$q(1-\alpha R_b^2 \sec^2\chi_b) = \sqrt{\alpha}R_b(1-q^2)(1-\alpha R_b^2)^{1/2}\tan\chi_b. \tag{36}$$

When $B > B_m$ the matching is possible only if the exterior solution is incoming.

Just as in the classical case the effective potential barrier in the Schwarzschild domain exists. But unlike the classical case, instead of the total reflection of the neutrino from the barrier there is a very small tunnelling effect.

8. Conclusion

The behaviour of zero mass particles is studied on the background of geometry pertaining to the non-rotating spherical collapsing object. A detailed analysis of null geodesics is made on the background of the Friedmann geometry matched on to the Schwarzschild geometry. Interesting phenomena such as backward emission, confinement of particles etc., are observed. Assuming a simple type of emission process, the flux

profiles at infinity are discussed. Effects such as bursts, discontinuities and decays are observed in flux profiles at various stages of the collapse.

Finally the massless Dirac equation pertaining to the neutrino studied on the background of these spacetimes. The examination of the solutions of the Dirac equation gives considerable insight into the propagation of neutrinos on the background geometry.

References

Dhurandhar S V and Vishveshwara C V 1981 *Astrophys. J.* **245** 1094
Dhurandhar S V and Vishveshwara C V 1983 (to be published)
Iyer B R, Dhurandhar S V and Vishveshwara 1982 *Phys. Rev.* **D25** 2053 (to be published)
Jaffe J 1969 *Ann. Phys.* **55** 374
Kembhavi A K and Vishveshwara C V 1980 *Phys. Rev.* **D22** 2349

Discussion

P. Majumdar: How are your conclusions affected if the neutrino has a small non-zero mass?

S. V. Dhurandhar: In our considerations the neutrinos have very high energies and therefore even if the neutrinos had a small non-zero mass they would be highly relativistic. Although such neutrinos would travel along timelike geodesics, the timelike geodesics would be very close to the null geodesics of which we have made detailed analysis. Therefore all the effects which are borne out by our investigations about the massless neutrino will essentially remain unaltered provided the rest-mass energy of the massive neutrino is small compared with its total energy.

§ III. ACCRETION DISK DYNAMICS

INTRODUCTION

One of the outstanding features of astrophysics today is the understanding of high energy emission mechanisms, particularly for objects like quasars and x-ray binaries. It is possible that accretion of matter onto compact objects could be the most effective process that lead to the release of a good fraction of the rest mass energy in an efficient way. In the case of accretion onto black holes 'accretion disks' that form around the black hole would be the agency for plasma processes that could emit high energy radiation. Hence the study of the dynamics of accretion disks forms a very important part of astrophysics, wherein one would consider the detailed structure and stability of rotating disks around black holes under the influence of gravitational, electromagnetic, centrifugal and pressure gradient forces.

In the following two articles such a study is presented for the dynamics of rotating (i) thin pressureless charged fluid disk confined to the equatorial plane and (ii) a thick uncharged perfect fluid disk with non zero pressure both the disks considered on the background of the Schwarzschild space-time. As has been pointed out it is very important in general to consider the dynamics of accretion disks under the influence of gravitational as well as self consistent electromagnetic fields. Also as the space-time outside a black hole would be highly curved, and for efficient radiation emission r_i, should be close to the event horizon, it is necessary that the entire discussion of the dynamics be carried out in the framework of general relativity. Though several attempts do exist with some of these aspects, it is important to note that there are still several open problems, some of which in our opinion, may be described as follows:

(i) the study of thick structured plasma disks with the inclusion of self generated as well as external magnetic fields.

(ii) the study of stability under axisymmetric perturbations with the inclusion of dissipative forces like radiation pressure, viscosity, convection, etc.

(iii) the study of passage of electromagnetic waves in the disks which would cause perturbations as well as the discussion of the frequency modulation and polarisation of the waves due to their passage in the dense, magnetised plasma of the disk.

One of the most interesting developments that would help in tackling these problems is the formulation of electrodynamics on curved space-time I due to Kip Thorne *et al* (1981) in a language similar to the nonrelativistic electrodynamics using a 3 + 1 slicing of the space-time and writing all the equations in terms of E and B fields unlike the usual covariant tensor notation. With the availability of this language one could hope for a better interaction between plasma astrophysicists and general relativists so that some of the outstanding problems in understanding high energy cosmic emission through plasma processes on curved space-time around black holes, may be solved.

Accretion disks around compact objects with self-consistent electromagnetic fields

A R PRASANNA
Physical Research Laboratory, Navarangpura, Ahmedabad 380 009, India

1. Introduction

Accretion is synonymous to gravitation, as the force of gravitation is always attractive. It is well known that a lump of matter left to itself will start contracting and when it gets compact will continue to accrete matter from its surroundings. An isolated compact object will accrete interstellar matter almost radially whereas the massive component of a binary will accrete matter from its companion in a preferred direction.

Such accretion leads to energy release which may be observed in the form of increased luminosity of the compact object, as given by

$$L = \tfrac{1}{2}\dot{M}V_{\mathrm{ff}}^{2}, \tag{1}$$

wherein V_{ff} is the free fall velocity of the accreting material. This velocity for a particle impinging upon the surface of a star may be obtained by equating its kinetic and potential energies which for a star of mass M and radius R is given by $V_{\mathrm{ff}} = \sqrt{2GM/R}$. Hence the increased luminosity turns out to be $\eta\dot{M}C^2$, wherein $\eta = GM/RC^2$ is the efficiency of energy release. This efficiency factor is $\sim 10^{-4}$ for a white dwarf ($M \sim 1\ M_0$, $R \sim 10^9$ cm) and $\eta \approx 0.14$ for a neutron star ($M \sim 1\ M_0$, $R \sim 10^6$ cm). If the compact object is a black hole then though its radius may be taken as to be $2GM/C^2$, we cannot take $\eta \approx 0{\cdot}5$, as black hole offers no hard surface for the particle to impinge upon. In this case the energy release through accretion comes mainly out of thermalisation of the matter in the strong potential well of the black hole, which is released before the matter is sucked in by the hole.

If the accreting matter has no angular momentum, the radiated energy would essentially come from the $\langle p dv \rangle$ work done in compressing the gas and if the accretion rate is below a certain critical value, only a small fraction of the energy will be radiated depending on the cooling time as compared to the characteristic infall time scale. If bremmstrahlung is the only cooling agency this critical value is given by (Lightman *et al* 1978)

$$\dot{M}_c \lesssim 10^{-8}(M/M_\odot)(T_\gamma/10^4)^{1/2} M_\odot\ \mathrm{yr}^{-1}. \tag{2}$$

In the case of isolated medium mass black holes in interstellar space this critical value is possibly attained, rendering them a weak source of x-rays.

On the other hand if the infalling matter possess angular momentum, as for example, in the case of binary systems, the accretion rate would be higher as well as the matter instead of falling radially, would form a rotating disk around the compact object. Once such a disk is formed, the two dominant processes of radiation from the disk seem to be

(i) bremmstrahlung and (ii) comptonisation. With the formation of the disk the gas cannot move radially without losing angular momentum. In order to transport angular momentum one invokes viscosity, which will ensure the process of mass loss to be very slow. This in turn would give sufficient time for thermalisation of the gas to temperatures of the order 10^8 °K rendering the disk to become a source of x-rays. Further due to scattering by relativistic electrons (inverse compton effect) one could get hard x-rays or by cold electrons (compton effect) one could get lower frequency emissions. Because of these possibilities, the accretion disk models have been put forward for a wide variety of objects like quasars, galactic x-ray sources, bursters, active nucleii of galaxies, elliptical galaxies, and also for the most enigmatic object SS433.

One of the early models considered for compact x-ray sources based on accretion disks was due to Pringle and Rees (1972) and for quasars due to Lynden Bell (1969). Subsequently several models have been suggested, a good review of which may be found in Lightman *et al* (1978). As mentioned earlier all these models come under the category of standard accretion disk models (SADM) which mainly are geometrically thin ($h \ll R$) and have three distinct regions (Shakura and Sunyaev 1976)

(i) inner region $p_r \gg p_g$, $\lambda^{es} \gg \lambda^{ff}$

(ii) middle region $p_r \ll p_g$, $\lambda^{es} \gg \lambda^{ff}$

(iii) outer region $p_r \ll p_g$, $\lambda^{es} \ll \lambda^{ff}$.

wherein p_r and p_g denote the radiation and gas pressure and λ^{es} and λ^{ff} denote the opacity due to electron scattering and free-free absorption. Further, the inner region is normally taken to be the hottest with maximum output of radiation. The energy production rate Q^+ is zero at $r = r_i$, the inner edge and reaches maximum at $r = (49/36)\,\mathrm{m}$ and decreases as r^{-3} for large r. The total luminosity,

$$L = 4\pi \int_{r_i}^{\infty} Q^+ r\,\mathrm{d}r = \tfrac{1}{2}\dot{M}_d Mc/r_i, \tag{3}$$

is independent of the nature of the dissipative forces and depends only on the accretion rate $\dot{M}_d$ and the inner radius r_i. Hence smaller the r_i, larger will be the luminosity for the same accretion rate.

The dynamics of the disk as envisaged in the standard disk models is governed by the laws of conservations of mass, angular momentum, energy and vertical momentum, nature of viscosity and by the law of radiative transfer, from inside of the disk to its upper and lower faces. The gas is assumed to be supported against the gravitational pull of the central compact object mainly by the rotation with Keplerian azimuthal velocity. Due to viscous forces the gas loses angular momentum and acquires radial velocity, while in the vertical direction the velocity is assumed to be subsonic so that the vertical structure is governed by the law of hydrostatic balance. Turbulence and small scale magnetic field contributes to the viscosity and one writes for the case of Keplerian motion, the integrated viscous stress $t_{r\varphi} = 2\alpha ph$, p = pressure, h = vertical height and α is a constant less than 1. The energy produced due to the friction is transferred to its surfaces. The medium is considered as optically thick with the opacity due to Thomson scattering and due to free-free absorption. To discuss the stationary state one also needs an equation of state with the total pressure due to p_r and p_g, and an equation connecting the radiation density to the thermodynamical properties of the gas. Starting with such steady state disks, various authors have considered the stability under varying kinds of perturbations.

Considering the disk dynamics we find that the most important feature missing is the dynamical role of the magnetic field. In fact as pointed out by Lightman *et al* (1978) at the high temperatures attained close to the black hole ($T \sim 10^{12}$ °K) the particle mean free paths are so long that a fluid dynamical treatment is not really self-consistent, unless collective effects are operative. However, the only saving grace is the magnetic field which if dynamically important at the accretion radius would affect the character of the flow. Even if it is initially negligible (like interstellar field) the field lines will be stretched radially during the inflow as a consequence of which the magnetic energy density will vary as r^{-4} and will become dynamically important. The effects of such fields was considered by Bisnovatyi–Kogan and coworkers, a detailed treatment of which may be found in Bisnovatyi–Kogan (1979).

However, another important contribution to the magnetic field would come from the disk itself wherein due to the rotation of the charged constituents of the gas current loops will be set up and these loops will produce a magnetic field. Thus it is our contention that in any discussion of the dynamics of accretion disks one should consider the effects due to self-generated electromagnetic fields. This would mean that when one considers the general conservation laws one should consider the effects of a self-generated electromagnetic field due to the motion of the fluid in the disk. Particularly if one is studying the disk around a black hole as the space-time curvature will be very strong in its vicinity the electromagnetic field so generated will be enhanced and thus will be dynamically very important. In the presence of magnetic field, as the charged particles can get really close to the black hole without being sucked in, the inner radius of the disk can be very close to the black hole and thus the general relativistic effects will be significant.

2. Equations of structure

To treat such a system as mentioned above, we will now consider the complete set of dynamical equations that govern the disk with the following assumptions:

(i) The disk is not massive in comparison with the black hole and as such the space-time structure supporting the disk is entirely governed by the black hole.

(ii) The electromagnetic field produced in and around the disk is also of small energy compared to the black hole mass that it does not affect the geometry, but itself gets modified by the geometry.

With these two assumptions our task now is to consider the dynamics of fluid disks (in general having both charge density and conductivity) rotating around a black hole. The disk is taken to be in equilibrium under the action of gravitational, centrifugal, pressure gradient and electromagnetic forces. Considering the disk to be made up of charged perfect fluid with pressure p, matter density ρ, charge density ε, conductivity σ, the laws of conservation of energy and momentum may be expressed through the covariant equations

$$T^j_{i;j} = 0, \tag{4}$$

with the covariant derivative taken w.r.t. the background metric and

$$T^j_i = (\rho + p/c^2) u_i u^j - p/c^2 \delta^j_i - \frac{4\pi}{c} E^j_i. \tag{5}$$

wherein the electromagnetic energy momentum tensor E_i^j is given by

$$E_i^j = F_{ik}F^{jk} - \tfrac{1}{4}\delta_i^j F_{kl}F^{kl}, \tag{6}$$

with the field tensor F_{ij} satisfying the covariant Maxwell's equations

$$F^{ij}_{;j} = -\frac{4\pi}{c}J^i, \quad F_{[ij,k]} = 0, \tag{7}$$

along with the generalised Ohm's law

$$J^i = c\varepsilon u^i + \sigma F^{ij}u_j \tag{8}$$

u^i being the fluid four-velocity. Considering the background geometry to be given by the metric

$$ds^2 = g_{ij}dx^i dx^j, \tag{9}$$

we will have the fluid four-velocity normalisation condition

$$g_{ij}u^i u^j = +1, \tag{10}$$

with

$$u^i = dx^i/ds. \tag{11}$$

With these we have the conservation equations (1) split up into the equation of continuity (mass conservation)

$$\rho_{;j}u^j + (\rho + p/c^2)u^j_{;j} = \frac{1}{c^3}F_{ik}J^k u^i, \tag{12}$$

and the momentum equations:

$$\left(\rho + \frac{p}{c^2}\right)u^i_{;j}u^j + \left(\frac{p}{c^2}\right)_{;j}(u^i u^j - g^{ij}) = \frac{1}{c^3}(F^i_k - F_{lk}u^i u^l)J^k. \tag{13}$$

Using the spatial 3-velocity V^α, defined through the relation $u^\alpha = u^0 V^\alpha/c$, these systems of equations may be written as

$$\begin{aligned}
&\left(\rho + \frac{p}{c^2}\right)[V^\alpha_{,\alpha} - \Gamma^0_{0\alpha}V^\alpha + c\Gamma^\alpha_{0\alpha} + \Gamma^\beta_{\alpha\beta}V^\alpha - \Gamma^0_{\alpha\beta}V^\alpha V^\beta/c] \\
&+\left(\frac{\partial\rho}{\partial t} + V^\alpha\frac{\partial\rho}{\partial x^\alpha}\right) - \frac{1}{c^2}\left(\frac{\partial p}{\partial t} + V^\alpha\frac{\partial p}{\partial x^\alpha}\right) + \frac{1}{c^2(u^0)^2}\left(g^{00}\frac{\partial p}{\partial t} + cg^{0\alpha}\frac{\partial p}{\partial x^\alpha}\right) \\
&+\frac{1}{c^2u^0}\Bigg[\varepsilon F^0_\alpha V^\alpha + \sigma\Bigg\{\left(F^0_\alpha F^\alpha_0 + \frac{1}{c}F^0_\alpha F^\alpha_\beta V^\beta\right) \\
&-2(u^0)^2\left(F_{0\alpha}F^\alpha_0 + \frac{2V^\alpha}{c}F_{0\beta}F^\beta_\alpha + \frac{V^\alpha V^\beta}{c}(F_{\alpha 0}F^0_\beta + F_{\alpha\gamma}F^\gamma_\beta)\right)\Bigg\}\Bigg] = 0
\end{aligned} \tag{14}$$

and

$$\begin{aligned}
&\left(\rho + \frac{p}{c^2}\right)(u^0)^2\Bigg[\frac{\partial V^\alpha}{\partial t} + V^\beta\frac{\partial V^\alpha}{\partial x^\beta} + c^2\left(\Gamma^\alpha_{00} - \frac{V^\alpha}{c}\Gamma^0_{00}\right) \\
&+2cV^\beta\left(\Gamma^\alpha_{0\beta} - \frac{V^\alpha}{c}\Gamma^0_{0\beta}\right) + V^\beta V^\gamma\left(\Gamma^\alpha_{\beta\gamma} - \frac{V^\alpha}{c}\Gamma^0_{\beta\gamma}\right)\Bigg]
\end{aligned}$$

$$= \left(g^{\alpha 0} - \frac{g^{00} V^{\alpha}}{c}\right)\frac{1}{c}\frac{\partial p}{\partial t} + \left(g^{\alpha\beta} - \frac{g^{0\beta} V^{\alpha}}{c}\right)\frac{\partial p}{\partial x^{\beta}}$$
$$+ \varepsilon u^{0}\left[F^{\alpha}_{0} + \frac{1}{c}F^{\alpha}_{\beta}V^{\beta} - \frac{1}{c^{2}}F^{0}_{\beta}V^{\alpha}V^{\beta}\right]$$
$$+ \frac{\sigma u^{0}}{c}\left[\left(F^{\alpha}_{i} - \frac{V^{\alpha}}{c}F^{0}_{i}\right)F^{i}_{0} + \frac{V^{\beta}}{c}F^{i}_{\beta}\left(F^{\alpha}_{i} - \frac{V^{\alpha}}{c}F^{0}_{i}\right)\right]. \tag{15}$$

These four equations (14) and (15) together with the set of eight Maxwell equations (7) constitute the system of equations governing the structure of a rotating charged fluid disk around a gravitating source which is in equilibrium under the influence of

(i) the gravitational field produced by the background geometry

(ii) the centrifugal force produced by the rotating disk (V^{α}) and

(iii) the self-consistent electromagnetic field produced by the moving charged fluid of the disk (F_{ij}).

However this system has only twelve equations whereas the number of unknowns are thirteen, *viz*., the velocity field V^{α}, the electromagnetic fields E^{α} and B^{α} and the fluid parameters ρ, p, ε and σ. Hence in order to close the system we need an equation of state with which the system will be completely determined.

It is clear from the general system of equations that it is extremely formidable to solve for the structure in general. Thus in the present analysis we look for the particular case (more to show the possible existence of a solution) of a geometrically thin disk of nonconducting ($\sigma = 0$) charged fluid with axisymmetric electromagnetic field having only the azimuthal component of velocity to be non-zero.

Taking the Schwarzschild geometry as the background space-time and rewriting the system of equations in terms of local Lorentz components of all field quantities V^{α}, F_{ij}, we get (Prasanna and Chakraborty 1980) for the steady state, the equations:

$$\left(\rho_0 + \frac{p_0}{c^2}\right)\left\{\frac{mc^2}{r^2} - \left(1 - \frac{2m}{r}\right)\frac{(V^{(\varphi)}_0)^2}{r}\right\} + \left(1 - \frac{(V^{(\varphi)}_0)^2}{c^2}\right)\left(1 - \frac{2m}{r}\right)\frac{\partial p_0}{\partial r}$$
$$= -\varepsilon_0\left(1 - \frac{2m}{r}\right)^{1/2}\left(1 - \frac{(V^{(\varphi)}_0)^2}{c^2}\right)^{1/2}\left[F_{0(r)(t)} + \frac{V^{(\varphi)}_0}{c}F_{0(r)(\varphi)}\right] \tag{16}$$

$$\left(\rho_0 + \frac{p_0}{c^2}\right)\frac{\cot\theta}{r}V^{(\varphi)^2}_0 = \left(1 - \frac{V^{(\varphi)^2}_0}{c^2}\right)\frac{1}{r}\frac{\partial p_0}{\partial\theta}$$
$$+ \varepsilon_0\left(1 - \frac{V^{(\varphi)^2}_0}{c^2}\right)^{1/2}\left[F_{0(\theta)(t)} + \frac{V^{(\varphi)}_0}{c}F_{0(\theta)(\varphi)}\right]. \tag{17}$$

$$\varepsilon_0\left(1 - \frac{V^{(\varphi)^2}_0}{c^2}\right)^{3/2}\left(1 - \frac{2m}{r}\right)^{1/2}F_{0(\varphi)(t)} = 0. \tag{18}$$

$$\frac{\partial}{\partial r}\left[r\left(1 - \frac{2m}{r}\right)^{1/2}F_{0(\theta)(t)}\right] + \frac{\partial}{\partial\theta}F_{0(t)(r)} = 0. \tag{19}$$

$$\frac{\partial}{\partial r}\left[r\left(1 - \frac{2m}{r}\right)^{1/2}F_{0(\varphi)(t)}\right] = 0. \tag{20}$$

$$\frac{\partial}{\partial\theta}[\sin\theta\, F_{(\theta)(\varphi)}] = 0 \tag{21}$$

$$\frac{\partial}{\partial r}[r^2 \sin\theta\, F_{0(\theta)(\varphi)}] + \frac{\partial}{\partial\theta}\left[r\sin\theta\left(1-\frac{2m}{r}\right)^{-1/2} F_{0(\varphi)(r)}\right] = 0, \tag{22}$$

$$\frac{\partial}{\partial\theta}[\sin\theta\, F_{0(r)(\theta)}] = 0, \tag{23}$$

$$\frac{\partial}{\partial r}\left[r\left(1-\frac{2m}{r}\right)^{1/2} F_{0(\theta)(r)}\right] = 0, \tag{24}$$

$$\frac{\partial}{\partial r}\left[r\left(1-\frac{2m}{r}\right)^{1/2} F_{0(\varphi)(r)}\right] + \frac{\partial}{\partial\theta}[F_{0(\varphi)(\theta)}] = -\frac{4\pi\varepsilon_0}{c}\left(1-\frac{V_0^{(\varphi)^2}}{c^2}\right), \tag{25}$$

$$\frac{\partial}{\partial r}[r^2 F_{0(t)(r)}] + \frac{r}{\sin\theta}\frac{\partial}{\partial\theta}\left[\sin\theta\left(1-\frac{2m}{r}\right)^{-1/2} F_{0(t)(\theta)}\right]$$
$$= 4\pi\varepsilon_0 r^2\left(1-\frac{2m}{r}\right)^{-1/2}\left(1-\frac{V_0^{(\varphi)^2}}{c^2}\right)^{-1/2} \tag{26}$$

The equation of continuity being identically satisfied through the requirements of conservation of charge and of mass explicitly.

3. Steady state solutions

As mentioned earlier our main aim is to solve the equations governing the structure of the disk self-consistently such that the electromagnetic field produced by the rotating disk equilibrates the centrifugal and pressure gradient forces of the disk along with the gravitational force of the central massive object. Looking at the set of equations (16) to (26), it is clear from (18) that

$$F_{0(\varphi)(t)} = 0, \tag{27}$$

i.e. the toroidal electric field is zero. This in fact is consistent with the Maxwell's equations (20) and (21) which are identically satisfied with (27), as well as the requirements of axisymmetry. Further it is clear from the equations that the toroidal magnetic field $F_{0(r)(\theta)}$, too may be taken to be zero as it is not coupled to any other physical parameter of the system and further is consistent with axisymmetry. This satisfies (23) and (24) identically and thus we will now be left with six equations (16), (17), (19), (22), (25) and (26), connecting the eight quantities ρ, p, $V_0^{(\varphi)}$, ε, E_r, E_θ, B_r and B_θ. Hence in order to solve this system we will need two more conditions. Though, in principle, one should choose an equation of state connecting ρ and p, prudence points out that it is better at this stage to solve for the electromagnetic field first and then look for consistent solutions.

Considering (22) and (25) governing the magnetic fields it is clear that by choosing a proper source function the magnetic field could be determined completely. If the right side of (25) is chosen as given then we have two possible cases.

(i) ε_0 is constant and u^φ is constant meaning that the charge density observed by the comoving observer is constant and the disk is having a differential rotation

(ii) u^φ/u^0 is constant indicating rigid rotation, and $b = \varepsilon_0(1-2m/r)^{-1/2}(1-V_0^{(\varphi)^2}/c^2)^{-1/2}$, the charge density as observed by the observer at infinity is constant.

We shall consider these two cases separately looking for complete solution of the structure equations.

3.1 *Differential rotation*

$u^{\varphi} = \Omega/c$ where Ω is a constant, gives the azimuthal angular velocity $V_0^{(\varphi)} = r\Omega \sin\theta\,(1 + r^2\Omega^2 \sin^2\theta/c^2)^{-1/2}$ and this together with ε_0 = constant would make (25) to be

$$\frac{\partial}{\partial r}\left[r\left(1-\frac{2m}{r}\right)^{1/2} F_{0(\varphi)(r)}\right]+\frac{\partial}{\partial\theta}F_{0(\varphi)(\theta)} = -\frac{4\pi\varepsilon_0}{c}r^2\Omega\sin\theta. \tag{28}$$

Considering the source of the magnetic field which is $\varepsilon_0\Omega$ it is obvious that the current loops are produced due to the motion of the charges in the disk and as such the magnetic field at infinity would be dipolar in character. Thus we can assume the magnetic field to be such that its components are given by

$$\begin{aligned} F_{0(\varphi)(\theta)} &= B_r = (2\mu\cos\theta/r^3)f(r),\\ F_{0(r)(\varphi)} &= B_\theta = (\mu\sin\theta/r^3)g(r), \end{aligned} \tag{29}$$

wherein $f(r)$ and $g(r)$ characterise the contributions due to the gravitational field and would go to 1 as $r \to \infty$, and μ is the dipole moment.

Substituting from (29) in (22) and (28) we get the equations for f and g to be

$$\frac{d}{dr}\left(\frac{f}{r}\right)+\left(\frac{g}{r^2}\right)\left(1-\frac{2m}{r}\right)^{-1/2} = 0, \tag{30}$$

$$\frac{d}{dr}\left[\frac{g}{r^2}\left(1-\frac{2m}{r}\right)^{1/2}\right]+\frac{2f}{r^3} = \frac{4\pi\Omega\varepsilon_0}{\mu c}r^2. \tag{31}$$

It is clear that in the absence of charges (*i.e.* outside the disk) where $\varepsilon_0 = 0$ these equations are same as the ones obtained by Ginzburg and Ozernoi (1965) for the external field of a magnetic star. Thus we look for a solution of (30) and (31) such that at the outer boundary of the disk, these solutions match with the solution of Ginzburg and Ozernoi. Hence we have the solutions for f and g to be

$$f = C_1 f_1 + C_2 f_2 - (2\pi\Omega\varepsilon_0/5\mu c)h_1 r^5, \tag{32}$$

with

$$f_1 = \left(\frac{r^3}{m^3}\right)\left[\ln\left(1-\frac{2m}{r}\right)+\left(\frac{2m}{r}\right)\left(1+\frac{m}{r}\right)\right],$$

$$f_2 = \frac{r^3}{2m^3},$$

$$h_1 = 1+\frac{4m}{r}+\left(\frac{8m^2}{r^2}\right)\ln\left(\frac{r}{2m}-1\right), \tag{33}$$

and

$$g = -C_1 g_1 - C_2 g_2 + \left(\frac{8\pi\Omega\varepsilon_0}{5\mu c}\right)h_2 r^5,$$

$$g_1 = \left(\frac{2r^2}{m^2}\right)\left[\frac{r}{m}\ln\left(1-\frac{2m}{r}\right)+\left(1-\frac{2m}{r}\right)^{-1}+1\right]\left(1-\frac{2m}{r}\right)^{1/2},$$

$$g_2 = \left(\frac{r^3}{m^3}\right)\left(1-\frac{2m}{r}\right)^{1/2},$$

$$h_2 = \left[1+\frac{3m}{r}+\frac{4m^2}{r^2}\ln\left(\frac{r}{2m}-1\right)+\frac{m}{r}\left(\frac{r}{2m}-1\right)^{-1}\right]\left(1-\frac{2m}{r}\right)^{1/2} \tag{34}$$

As there are two constants C_1 and C_2 we need to have two conditions and we choose them to be the continuity of fields at the outer and inner edge of the disk. As pointed out at the outer edge the fields match the Ginzburg–Ozernoi solution whereas at the inner edge we need to find the solution for the region between the inner edge and the stellar surface with $\varepsilon_0 = 0$.

In the absence of gravitational field the magnetic field inside a circular current loop of radius a is given by

$$F_{(r)(\varphi)} = \hat{B}_\theta = -\mu \sin\theta/a^3,\ F_{(\varphi)(\theta)} = \hat{B}_r = \mu\cos\theta/a^3. \tag{35}$$

Generalising this to the case of the disk and introducing corrections due to space-time curvature we have

$$\hat{B}_\theta = -\mu\hat{D}\sin\theta\,\hat{g}(r),\ \hat{B}_r = \mu\hat{D}\cos\theta\,\hat{f}(r),$$

with

$$\hat{D} = 8(R_b - R_a)/m^3\,(R_b^4 - R_a^4). \tag{36}$$

We determine $\hat{f}$ and $\hat{g}$ by solving the Maxwell's equations on the background geometry and the solution so obtained is given by

$$\hat{f} = \hat{C}_1\left[\ln\left(1-\frac{2m}{r}\right) + \frac{2m}{r}\left(1+\frac{m}{r}\right)\right] + \hat{C}_2, \tag{37}$$

$$\hat{g} = \left\{\hat{C}_1\left[\ln\left(1-\frac{2m}{r}\right) + \frac{m}{r} + \frac{m}{r}\left(1-\frac{2m}{r}\right)^{-1}\right] + \hat{C}_2\right\}\left(1-\frac{2m}{r}\right)^{1/2} \tag{38}$$

As the fields should be regular at $r = 2m$, we choose $\hat{C}_1 = 0$ and to get the flat space components as $m \to 0$, we choose $\hat{C}_2 = 1$. Thus the field components for the region $r < r_a$ are given by

$$\hat{B}_r = \mu\hat{D}\cos\theta,\ \hat{B}_\theta = -\mu\hat{D}\sin\theta\left(1-\frac{2m}{r}\right)^{1/2} \tag{39}$$

Thus at $r = r_b$ the B_r and B_θ coincide with the Ginzburg–Ozernoi solution whereas at $r = r_a$ they coincide with $\hat{B}_r$ and $\hat{B}_\theta$ respectively. These boundary conditions can be used to determine the constants of integration and thus the magnetic field may be determined uniquely.

In order to obtain the electric field structure we now reconsider the set of equations governing the steady state and make use of the fact that the electromagnetic field must be self-consistent with the charge distribution and the velocity field. Hence from (16) and (17) we can solve for the components $E_r\,(= F_{0(r)(t)})$ and $E_\theta\,(= F_{0(\theta)(t)})$ algebraically in terms of p_0, ρ_0, B_r and B_θ. Using these values of E_r and E_θ in (19) and (26) the system may be closed giving different conditions connecting p_0, ρ_0 and the magnetic field. In this way we can solve the entire set of steady state equations self-consistently and obtain the components of all the physical parameters in terms of ε_0 and Ω. In principle when we solve for p_0 and ρ_0 first order differential equations we will have two more constants of integration which need to be specified. In order to get these we use again the boundary condition of continuity of the electric field.

As the field structure of the magnetic field outside the disk has been assumed to be dipolar it is clear that we should take for the electric field only a monopole structure outside the distribution. Thus if we now look at the structure of the electric field inside and outside a charged current loop we know that the field inside is zero whereas outside

there will only be the E_r component non-zero given by the Coulomb field q/r^2, q being the total charge on the ring. Generalising this to the field outside a charged disk we expect the field to be for $R < R_a$

$$E_r = 0, \quad E_\theta = 0, \tag{40}$$

whereas for $R > R_b$

$$E_\theta = 0,$$

$$E_r = q/r^2, \quad q = 2\pi\bar{\varepsilon}_0 m^2 \int_{R_a}^{R_b} (1 + R^2\omega^2)^{1/2} R \, \mathrm{d}R. \tag{41}$$

Though in principle this way one can determine all the steady state parameters uniquely, in practice it is quite formidable to solve, as the differential equations for ρ_0 and p_0 are very complicated. (Equations (4.19) and (4.20) of Prasanna and Chakraborty (1980)). Hence we make one more simplifying assumption which is reasonable. We shall consider the case of a geometrically thin disk confined to the equatorial plane $\theta = \pi/2$. As the hydrostatic pressure p_0 is supposed to give the vertical balance of the disk against self-gravitation, in the case of infinitesimally thin disk without self-gravitation, p_0 can be assumed to be zero as there is no vertical structure. Thus using $\theta = \pi/2$ and $p_0 = 0$, the components of the electromagnetic field turn out to be

$$B_r = 0, \quad E_\theta = 0, \quad B_\theta = (\mu/r^3) g(r),$$

$$E_r = \left[\frac{\rho_0 c^2}{\varepsilon_0 r}\left\{\frac{m}{r} - \left(1 - \frac{3m}{r}\right)\frac{r^2\Omega^2}{c^2}\right\}\left(1 - \frac{2m}{r}\right)^{-1/2} + \frac{\Omega\mu}{cr^2} g\right]\left(1 + \frac{r^2\Omega^2}{c^2}\right)^{-1/2}. \tag{42}$$

In order to solve for ρ_0, we first introduce the height coordinate $h = R_m(\pi/2 - \theta)$ and two densities $\bar{\rho}_0$, $\bar{\varepsilon}_0$, $\bar{\alpha}$ and $\bar{\beta}$ as given by equation (4.20) of Prasanna and Chakraborty (1980)

$$\rho_0 = \bar{\rho}_0(R)\delta(h), \; \alpha = \bar{\alpha}\delta(h), \; \beta = \bar{\beta}\delta(h), \; \varepsilon_0 = \bar{\varepsilon}_0\delta(h) \tag{43}$$

(in equation (4.20) of Prasanna and Chakraborty (1980)) and integrate with respect to h to get the equation

$$\frac{\mathrm{d}}{\mathrm{d}R}\left[\bar{\rho}_0 R\left\{\frac{1}{R} - \left(1 - \frac{3}{R}\right)R^2\omega^2\right\}\left(1 - \frac{2}{R}\right)^{-1/2}(1 + R^2\omega^2)^{-1/2}\right.$$
$$\left. + \omega(1 + R^2\omega^2)^{-1/2}\left\{\bar{\alpha}(-C_1 g_1 - C_2 g_2) + \frac{8\pi}{5}\omega R^5 h_2 \bar{\beta}\right\}\right]$$
$$= 4\pi\bar{\beta}R^2\left(1 - \frac{2}{R}\right)^{1/2}(1 + R^2\omega^2)^{1/2}. \tag{44}$$

Expanding the factor $(1 - 2/R)^{1/2}$ and considering terms up to $1/R^6$, we can integrate this equation and get the solution for $\bar{\rho}_0$.

$$\bar{\rho}_0 = \frac{1}{R}\left(1 - \frac{2}{R}\right)^{1/2}(1 + R^2\omega^2)^{1/2}\left\{\frac{1}{R} - \left(1 - \frac{3}{R}\right)R^2\omega^2\right\}^{-1}$$
$$\left[4\pi\bar{\beta}D - \omega(1 + R^2\omega^2)^{-1/2}\left\{\bar{\alpha}(-C_1 g_1 - C_2 g_2) + \frac{8\pi}{5}\omega R^5 h_2 \bar{\beta}\right\} + C_3\right]. \tag{45}$$

wherein

$$D = (1 + R^2\omega^2)^{3/2}\left(\frac{R}{4}\omega^2 - \frac{231}{48R^3} + \frac{1}{3\omega^2}\right)$$
$$+ (1 + R^2\omega^2)^{1/2}\left(\frac{5}{2} - \frac{35}{8R} - \frac{63}{16R^2} + R\left(\frac{3}{4} - \frac{1}{8\omega^2}\right)\right)$$
$$+ \ln\left[R\omega + (1 + R^2\omega^2)^{1/2}\right]\left(41 - \frac{1}{\omega^2}\right)\Big/ 8\omega$$
$$+ \left(\frac{5}{4} + \frac{63\omega^2}{32}\right)\ln\left[\frac{(1 + R^2\omega^2)^{1/2} - 1}{(1 + R^2\omega^2)^{1/2} + 1}\right]. \tag{46}$$

From this expression for $\bar{\rho}_0$ it is obvious that it has singularities at R given by

$$(R-2)^{1/2}\,(R^3\omega^2 - 3R^2\omega^2 - 1) = 0. \tag{47}$$

At $R = 2$ the singularity is due to 'event horizon' of the back ground space and as $R_a \gg 2$ we do not have to worry about it. On the other hand if we take

$$R^3\omega^2 - 3R^2\omega^2 - 1 = 0, \tag{48}$$

we get

$$R_s = 1 + \frac{1}{(2\omega)^{1/3}}\left[\{1 + 2\omega^2 + (1 + 4\omega^2)^{1/2}\}^{1/3}\right.$$
$$\left. + \{1 + 2\omega^2 - (1 + 4\omega^2)^{1/2}\}^{1/3}\right], \tag{49}$$

which for lower values of ω lies very far away and as ω increases approaches the value 3 asymptotically. But if we look at the structure of the electric and magnetic fields outside the disk, the monopole electric field and the dipole magnetic field are just the first order terms of the respective fields in multipole expansion which obviously means that we have already restricted the velocities of the charged particles to be small compared to c. Thus we must for consistency have only the slowly rotating disks with ω small, which means that R_s would lie very far away. Obviously given an ω we shall choose $R_b \ll R_s$ thus ensuring the density function regular in its domain of definition.

Before considering the profiles we shall first evaluate the constants of integration C_1, C_2 and C_3. However, as we will have four boundary conditions—continuity of E_r and B_θ at R_a and R_b, we can evaluate one more parameter and thus we shall calculate $\bar{\beta}$ the constant associated with the charge density $\bar{\varepsilon}_0$. Introducing dimensionless parameters we can write

$$x = (m\bar{\varepsilon}_0/c^2)(B_\theta)_{\pi/2} = \frac{\bar{\alpha}}{R^3}(-C_1 g_1 - C_2 g_2) + \frac{8\pi}{5}\bar{\beta}\omega h_2 R^2,$$

$$y = \left(\frac{m\bar{\varepsilon}_0}{c^2}\right)(E_r)_{\pi/2} = \frac{4\pi\bar{\beta}D}{R^2} + \frac{C_3}{R^2},$$

$$\bar{\rho}_0 = (y - V_0 x)R\,(1 + R^2\omega^2)^{1/2}\left(1 - \frac{2}{R}\right)^{1/2}\left\{\frac{1}{R} - \left(1 - \frac{3}{R}\right)R^2\omega^2\right\}^{-1},$$

$$V_0 = \frac{R\omega}{(1 + R^2\omega^2)^{1/2}},$$

and

$$C_1 = \frac{g_{2b}\left\{\hat{D}R_b^3\left(1-\frac{2}{R_b}\right)^{1/2}+A_1\right\}+g_{2a}\left\{\frac{3}{8}g_{1b}+A_2\right\}}{(g_{1a}g_{2b}-g_{1b}g_{2a})},$$

with

$$C_2 = \frac{g_{1b}\left\{\hat{D}R_a^3\left(1-\frac{2}{R_a}\right)^{1/2}+A_1\right\}+g_{1a}\left\{\frac{3}{8}g_{1b}+A_2\right\}}{(g_{2a}g_{1b}-g_{2b}g_{1a})},$$

$$\bar{\beta} = \frac{Am\bar{\varepsilon}_0^2}{2c^2(D_b-D_a)}, \quad C_3 = \frac{-2\pi AD_a m\bar{\varepsilon}_0^2}{c^2(D_b-D_a)}$$

$$A_1 = \frac{16R_a^5 h_{2a}A}{5(D_b-D_a)(R_b^4-R_a^4)},$$

$$A_2 = \frac{-16R_b^5 h_{2b}A}{5(D_b-D_a)(R_b^4-R_a^4)},$$

$$A = \int_{r_a}^{r_b}\left(1+\frac{r^2\Omega^2}{c^2}\right)^{1/2} r\,dr,$$

wherein the subscripts a and b denote the values of the functions at r_a and r_b respectively.

Figures 1 and 2 denote the profiles of density electric and magnetic fields for different choice of the rotation parameter ω and for different sizes.

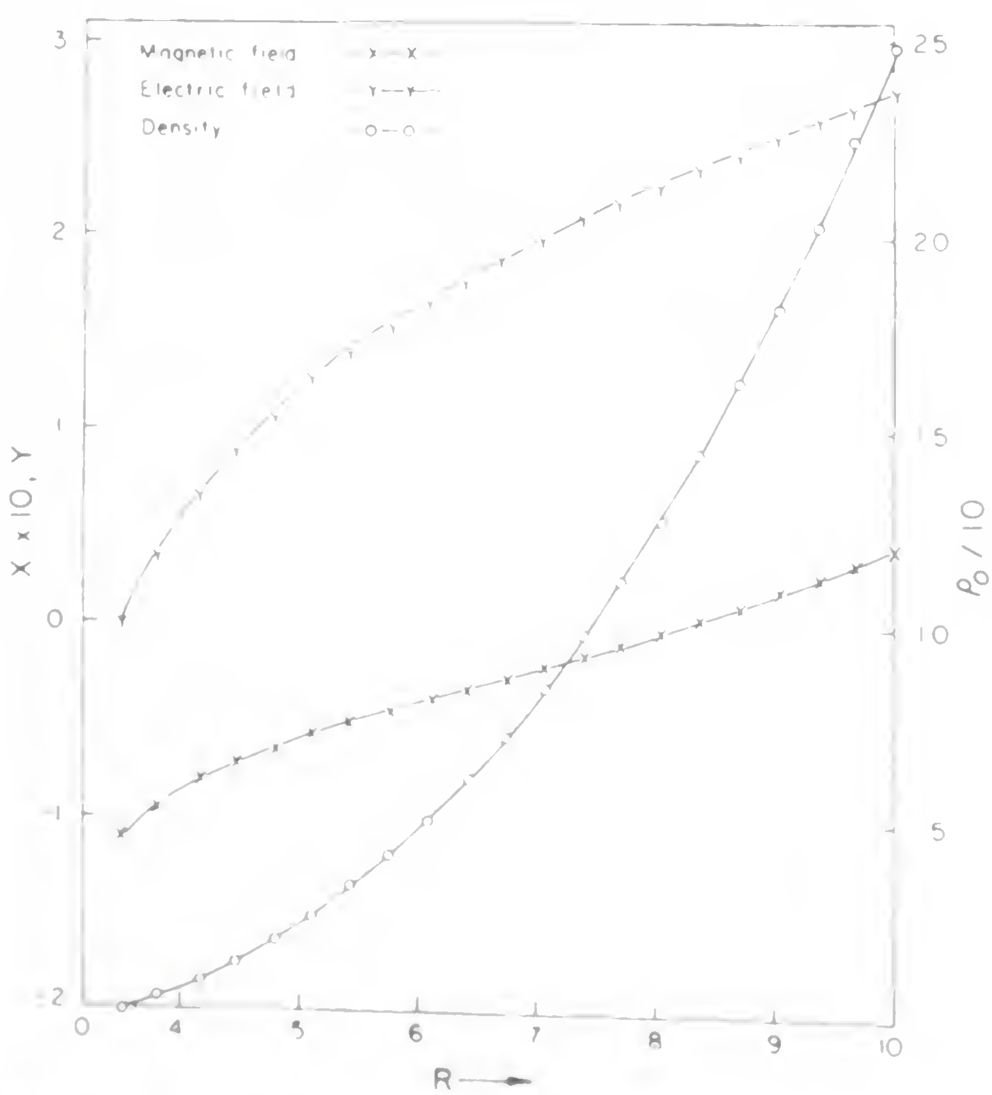

Figure 1. Differential Rotation $R_a = 3.5$, $R_b = 10$, $\omega = 0.004$

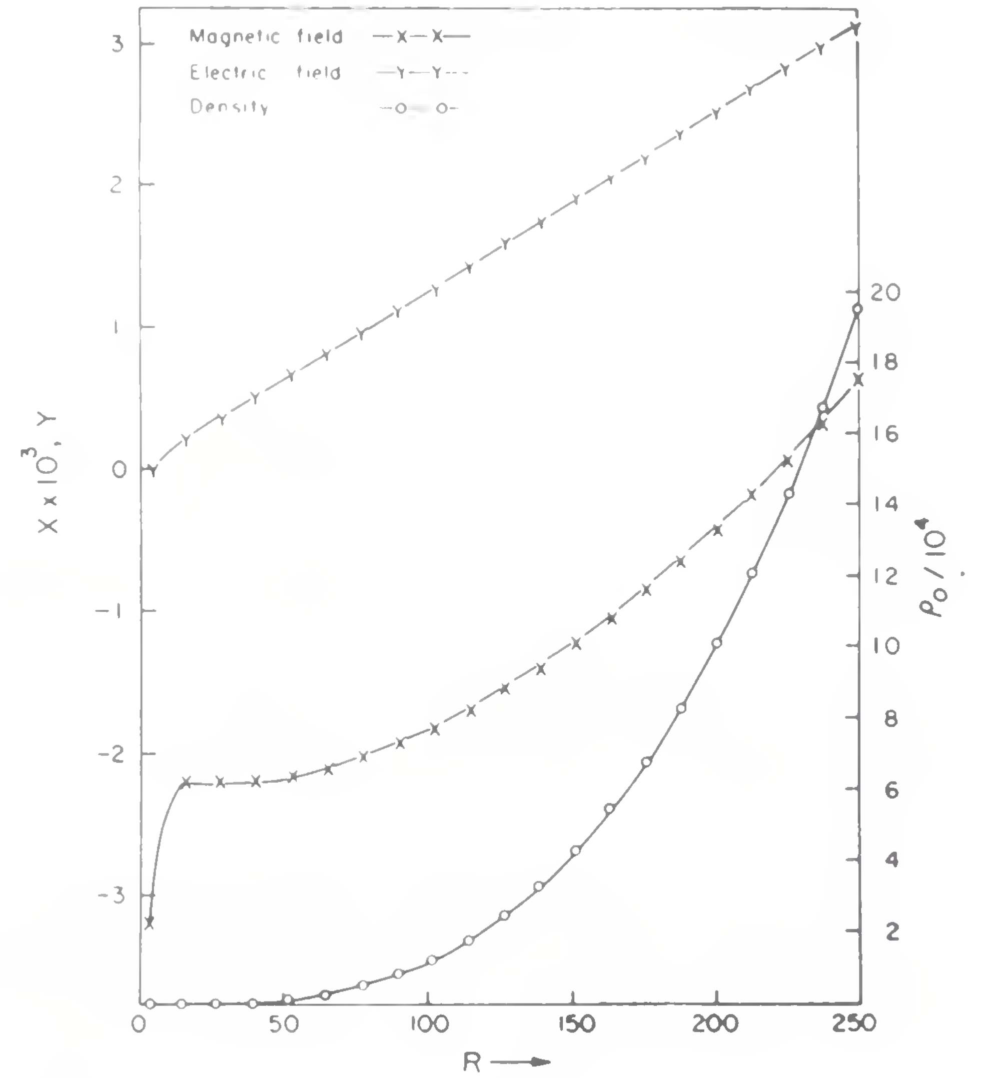

Figure 2. Differential Rotation $R_a = 3{\cdot}5$, $R_b = 250$, $\omega = 0{\cdot}3 \times 10^{-5}$

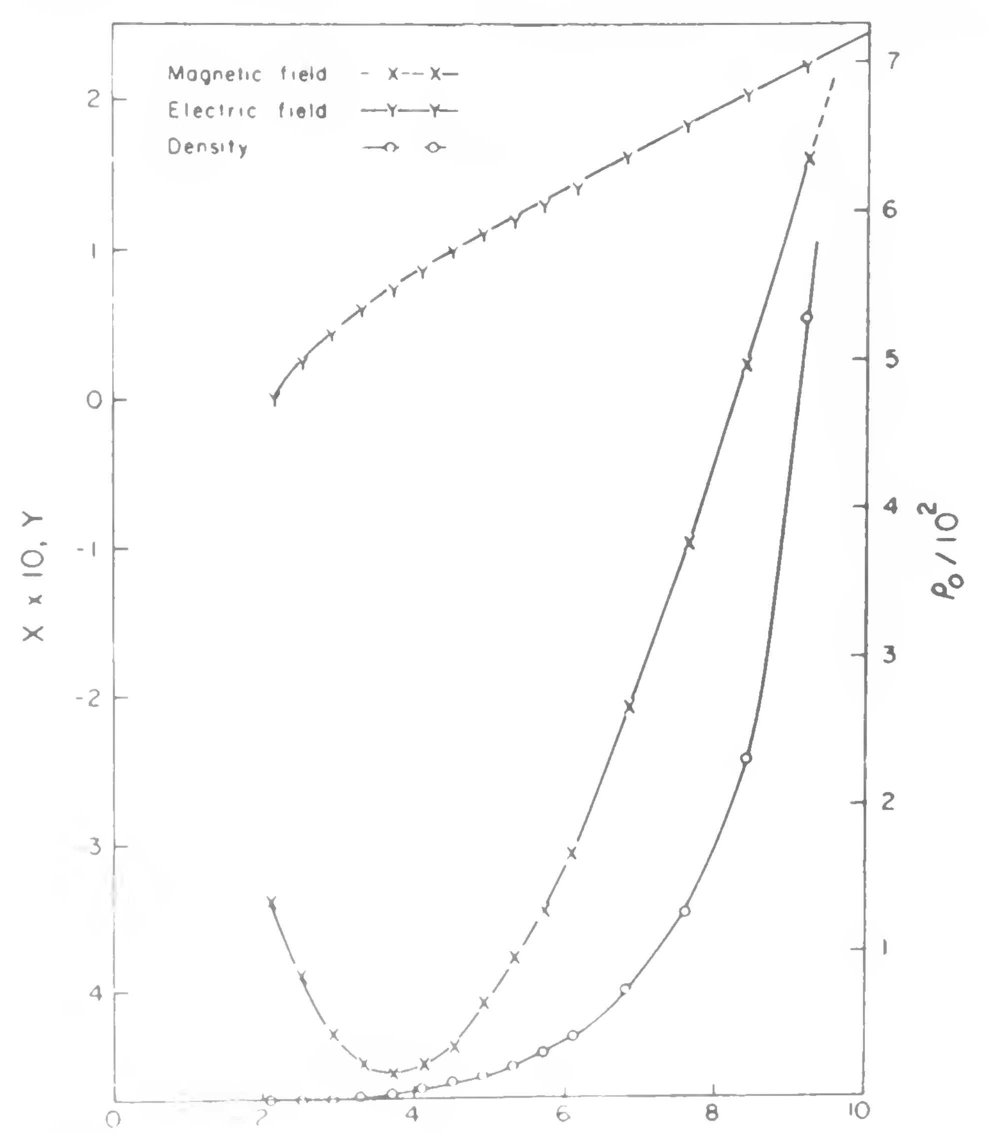

Figure 3. Rigid Rotation $R_a = 2{\cdot}1$, $R_b = 10$, $\omega = 0{\cdot}031$

3.2 Rigid rotation

As mentioned earlier, if we choose u^{φ}/u^0 as constant, we will have a rigidly rotating disk from the point of view of the external observer, and as such we appropriately take the charge density as observed by the far away observer $b = \varepsilon_0(1-2m/r)^{-1/2}(1-V_0^{(\varphi)^2}/c^2)^{-1/2}$, to be a constant. Taking $u^{\varphi}/u^0 = \Omega/c$, Ω a constant we get the three-velocity

$$V_0^{(\varphi)^2} = r\Omega \sin\theta\left(1-\frac{2m}{r}\right)^{-1/2}.$$

The equations governing a dipolar type of magnetic field are given by

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{f}{r}\right)+\left(\frac{g}{r^2}\right)\left(1-\frac{2m}{r}\right)^{-1/2} = 0$$

$$\frac{\mathrm{d}}{\mathrm{d}r}\left[\frac{g}{r^2}\left(1-\frac{2m}{r}\right)^{1/2}\right]+\left(\frac{2f}{r^3}\right) = \left(\frac{4\pi\Omega b}{\mu c}\right)r^2.$$

Obviously the solutions are the same as we had in (32) to (34) with ε replaced by b. Adopting a similar technique as was used in the case of differentially rotating disk we can get the physical quantities associated with a thin pressureless disk confined to the equatorial plane, to be

$$x = \frac{m\bar{b}}{c^2}(B_\theta)_{\pi/2} = \frac{\bar{\alpha}}{R^3}(-C_1 g_1 - C_2 g_2) + \frac{8\pi}{5}\bar{\beta}\omega R^2 h_2$$

$$y = \frac{m\bar{b}}{c^2}(E_r)_{\pi/2} = \frac{4\pi\bar{\beta}R}{3} + \frac{C_3}{R^2}$$

$$\rho_0 = [y - V_0 x]R\left(1-\frac{2}{R}-R^2\omega^2\right)\left(\frac{1}{R}-R^2\omega^2\right)^{-1}$$

with

$$V_0 = R\omega\left(1-\frac{2}{R}\right)^{-1/2}$$

$$C_1 = \frac{\left[g_{2b}\left\{\hat{D}R_a^3\left(1-\frac{2}{R_a}\right)^{1/2}+A_1\right\}+\left\{\frac{3}{8}g_{1b}+A_2\right\}\right]}{(g_{1a}g_{2b}-g_{1b}g_{2a})}$$

$$C_2 = \frac{\left[g_{1b}\left\{\hat{D}R_a^3\left(1-\frac{2}{R_a}\right)^{1/2}+A_1\right\}+g_{1a}\left\{\frac{3}{8}g_{1b}+A_2\right\}\right]}{(g_{2a}g_{1b}-g_{2b}g_{1a})}.$$

$$C_3 = \frac{-4\pi}{3}\bar{\beta}R_a^3,$$

$$\bar{\beta} = \frac{3}{2}\frac{Am\bar{b}^2}{c^2(R_b^3-R_a^3)},$$

$$A_1 = \frac{48R_a^5 h_{2a}A}{5(R_b^3-R_a^3)(R_b^4-R_a^4)}.$$

$$A_2 = \frac{-48 R_b^5 h_{2b} A}{5(R_b^3 - R_a^3)(R_b^4 - R_a^4)},$$

$$A = \int_{R_A}^{R_B} R\left(1 - \frac{2}{R}\right)^{1/2} \mathrm{d}R.$$

Figures 3 and 4 show the profiles of the density distribution and the fields for the rigidly rotating disks with different values of Ω.

4. Stability analysis

Having obtained the steady-state structure of pressureless thin disks for differential as well as rigid rotations we have also discussed the stability of such disks under purely radial perturbations. The general perturbation equations for charged fluid disks are as given in §3 of Prasanna and Chakraborty (1980), and for purely radial perturbations we shall have the pulsation equation as in (5.1) to (5.4) of the same reference. The self-adjoint characteristic equation may be obtained as

$$\frac{m^2\sigma^2}{c^2}\int_{R_a}^{R_b}\left(1 - \frac{2}{R}\right)^{-1}(-\bar{\rho}_0 Z + x)^2 y^2 \,\mathrm{d}R$$
$$= \int_{R_a}^{R_b}\left(1 - \frac{2}{R}\right)\left[\frac{\mathrm{d}}{\mathrm{d}R}\{(-\bar{\rho}_0 Z + x)y\}\right]^2 \mathrm{d}R$$

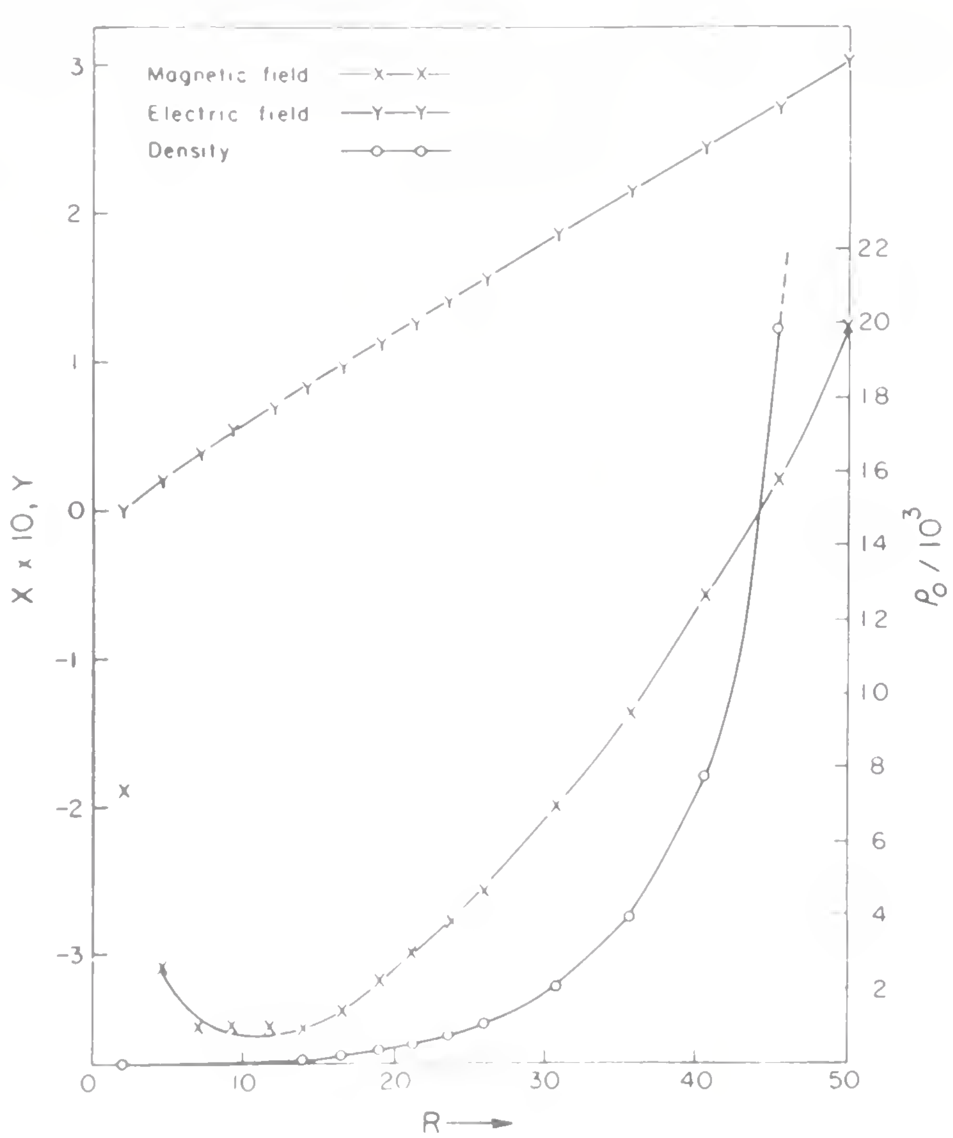

Figure 4. Rigid rotation $R = 2{\cdot}1$, $R = 50$, $\omega = 0{\cdot}003$

$$- 2\omega\pi\bar{\beta}\int_{R_a}^{R_b} y^2 \left\{ R\left(1-\frac{2}{R}\right)^{1/2} \frac{d}{dR}(-\bar{\rho}_0 Z + x) - (-\bar{\rho}_0 Z + x)\left(1-\frac{2}{R}\right)^{-1/2} F \right\} dR$$

wherein

$$F = \left\{ 2\left(1-\frac{2}{R}\right) R^2\omega^2 (1+R^2\omega^2)^{-1} + 1 - \frac{3}{R} \right\},$$

for the case of differential rotation and

$$F = \left(1 - \frac{1}{R}\right)$$

for rigid rotation. As all the quantities are known here except σ^2, we can evaluate by using a trial function of the form

$$y = (R - R_a)(R_b - R),$$

and choosing different sets of R_a, R_b and ω. It was found that in all the cases considered $\sigma^2 > 0$ *indicating the stability of these disks* under radial perturbation. This result is not altogether unexpected as we have restricted the disks to be slowly rotating and further did not have any dissipation of energy from the system.

It would indeed be very important to consider structured thick disks with non-zero pressure as then and only then one can allow for regimes with different densities and temperatures effecting physical processes leading to emission of radiation.

References

Bisnovatyi Kogan 1969 *Rev. del Nuovo Cimento* **2**

Ginzburg V L and Ozernoi I 1965 *Sov. Phys. JETP* **20** 689

Lightman A P, Shapiro S L and Rees M J 1978 *Physics and astrophysics of black holes and neutron stars* (eds) R Giacconi and R Ruffini (Amsterdam: North Holland)

Lynden Bell J 1969 *Nature (London)* **223** 690

Macdonald D and Thorne K S 1981 *Black hole electrodynamics* (March 1981, Orangeaid preprint, Caltech)

Prasanna A R and Chakraborty D K 1980 *J. Astrophys. Astron.* **2** 1

Pringle J and Rees M J 1972 *Astron. Astrophys.* **21** 1

Shakura N I and Sunyaev R A 1976 *Mon. Not. R. Astrophys. Soc.* **175** 613

Thorne K S and Macdonald D 1981 Electrodynamics in curved space-time: 3 + 1 formulation (Orangeaid preprint, Caltech)

Discussion

B R Iyer: Why was the field outside the disk matched to the Ginzburg Ozernoi dipole field

A R Prasanna: As the disk was considered to be slowly rotating with a small net charge for an observer at infinity the fields would be essentially a monopole electric field associated with a dipole magnetic field. Hence the matching with the Ginzburg Ozernoi field

Structure and stability of thick disks around black hole

D K CHAKRABORTY

Physical Research Laboratory, Navarangpura, Ahmedabad 380 009, India
*Permanent address: Department of Physics, Science College, Raipur 492 002, India

1. Introduction

Many astronomical objects are believed to radiate energy released by mass accretion onto a compact star: a neutron star or a black hole. The accreting matter forms a rotating disk if it has sufficient angular momentum. It is well known that the disk has a geometrically thick structure for a super-critical accretion rate (Shakura and Sunyaev 1973). On the other hand the disk is thin if the accretion rate is very low (subcritical accretion). Thin disk models were developed by Pringle and Rees (1972), Shakura and Sunyaev (1973) and Novikov and Thorne (1973), later termed as standard α-model. It was shown that the instabilities (Lightman 1974a, b; Shakura and Sunyaev 1976), probably cause the disk to puff off and to become geometrically thick.

Models of the thick accretion disk have been proposed by various authors (Wiita *et al* 1980; Paczynski and Wiita 1980; Jaroszynski *et al* 1980; Abramowicz *et al* 1980). The surface of the disk forms a double funnel around the rotation axis of the central compact star. The inner edge of the disk extends down to the marginally bound circular orbit r_{mb} ($= 4\,GM/c^2$ for a Schwarzschild black hole) and forms a cusp (Kozlowoski *et al* 1978; Fishbone and Moncrief 1976). All the studies of the thick disks made so far, concern only the steady state structure and the resultant luminosity. It is very important to study the stability of such disks.

Recently we (Chakraborty and Prasanna 1981, 1982) have analysed the structure and stability of thick fluid disk rotating around a compact star. The disk, we have considered, is made up of perfect fluid whose mass is negligible compared to that of the central star. The gravitational field is therefore solely determined by the central source which we have regarded as a non-rotating compact star. The disk is in equilibrium under the gravitational, centrifugal and pressure gradient forces. After obtaining a class of steady state solutions we have studied the stability of such disk under axi-symmetric perturbations. We use the variational principle as developed by Chandrasekhar and Friedman (1972a, b) to calculate the critical value of the adiabatic index for the neutral mode of deformation.

2. Steady state solutions

2.1 *The fundamental equations governing the dynamics*

The fundamental equations governing the dynamics of a non-self-gravitating perfect fluid disk rotating around a compact star can be derived from (Prasanna and Chakraborty 1981; Chakraborty and Prasanna 1982)

(i) the law of conservation of energy-momentum

$$T^{ij};j = 0, \tag{1}$$

where T^{ij} is the energy-momentum tensor given by

$$T^{ij} = pg^{ij} + (p + \rho c^2)u^i u^j, \tag{2}$$

wherein p, ρc^2 and u^i are pressure, energy density and for velocity respectively,

(ii) the laws of thermodynamics which include the law of conservation of baryon number

$$(nu^i);i = 0, \tag{3}$$

where n is the number density of baryons and the second law of thermodynamics

$$du + pdV = Tds, \tag{4}$$

where u, V, T and s are respectively the internal energy, volume, temperature and entropy. The above set of equations is supplemented by a suitable equation of state to make the system of equations close.

As u^i satisfies the normalisation relationship $u_i u^i = -1$, (1) may be resolved into the equation of continuity

$$\rho_{,i} u^i + (\rho + p/c^2)u^i;i = 0, \tag{5}$$

and the equation

$$(\rho + p/c^2)u^i;j\, u^j = -\frac{1}{c^2}(g^{ij} + u^i u^j)p_{,j}. \tag{6}$$

Using (3), (4), (5) and an equation of state

$$\rho c^2 = nM_0 c^2 + p/(\gamma - 1), \tag{7}$$

where M_0 is the rest mass of each baryon and $\gamma(= cp/c_v)$ is the adiabatic index, we obtain

$$\frac{u^j n^\gamma}{(\gamma - 1)} \frac{\partial}{\partial x^j}\left(\frac{p}{n^\gamma}\right) = 0, \tag{8}$$

expressing the conservation of entropy along the line of flow of the fluid. Equations of momentum conservation can be obtained from (6) which along with the continuity equation (5), baryon conservation equation (3) and equation (8) of adiabatic flow forms the complete set of equations governing the dynamics of the disk.

For the background geometry we adopt Schwarzschild metric

$$ds^2 = -\left(1 - \frac{2m}{r}\right)c^2 dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2). \tag{9}$$

In order to make easy comparison with the corresponding Newtonian case, we introduce 3-velocity v^α by $u^\alpha = v^\alpha u^0/c$ and further write the equations in terms of local Lorentz frame components defined by the orthonormal tetrad appropriate to the background metric. Thus the complete set of equations governing the dynamics of the disk is given by

(a) the momentum equations

$$(\rho + p/c^2)\left[\frac{Dv^{(r)}}{Dt} + \frac{mc^2}{r^2}\left(1 - \frac{v(r)^2}{c^2}\right) - \left(1 - \frac{2m}{r}\right)\left\{\frac{v(\theta)^2 + v(\varphi)^2}{r}\right\}\right]$$
$$= -\left(1 - \frac{v^2}{c^2}\right)\left[\left(1 - \frac{2m}{r}\right)\frac{\partial p}{\partial r} + \frac{v^{(r)}}{c^2}\frac{\partial p}{\partial t}\right], \tag{10}$$

$$\left(\rho + \frac{p}{c^2}\right)\left[\frac{Dv^{(\theta)}}{Dt} + \left(1 - \frac{3m}{r}\right)\frac{v^{(r)}v^{(\theta)}}{r} - \left(1 - \frac{2m}{r}\right)^{\frac{1}{2}}\frac{\cot\theta}{r}v(\varphi)^2\right]$$
$$= -\left(1 - \frac{v^2}{c^2}\right)\left[\left(1 - \frac{2m}{r}\right)^{\frac{1}{2}}\frac{1}{r}\frac{\partial p}{\partial \theta} + \frac{v^{(\theta)}}{c^2}\frac{\partial p}{\partial t}\right], \tag{11}$$

$$\left(\rho + \frac{p}{c^2}\right)\left[\frac{Dv^{(\varphi)}}{Dt} + \left(1 - \frac{3m}{r}\right)\frac{v^{(r)}v^{(\varphi)}}{r} + \left(1 - \frac{2m}{r}\right)^{\frac{1}{2}}\frac{\cot\theta}{r}v^{(\varphi)}v^{(\theta)}\right]$$
$$= -\left(1 - \frac{v^2}{c^2}\right)\left[\left(1 - \frac{2m}{r}\right)^{\frac{1}{2}}\frac{1}{r\sin\theta}\frac{\partial p}{\partial \varphi} + \frac{v^{(\varphi)}}{c^2}\frac{\partial p}{\partial t}\right]. \tag{12}$$

(b) the continuity equation

$$\left(\rho + \frac{p}{c^2}\right)\left(1 - \frac{2m}{r}\right)^{\frac{1}{2}}\left\{\left(1 - \frac{2m}{r}\right)^{\frac{1}{2}}\frac{1}{r^2}\frac{\partial}{\partial r}(r^2 v^{(r)}) + \frac{1}{r\sin\theta}\right.$$
$$\left.\times\left[\frac{\partial}{\partial\theta}(\sin\theta v^{(\theta)}) + \frac{\partial v^{(\varphi)}}{\partial\varphi}\right]\right\} + \frac{D}{Dt}\left(\rho - \frac{p}{c^2}\right) + \frac{1}{c^2}\left(1 - \frac{v^2}{c^2}\right)\frac{\partial p}{\partial t} = 0, \tag{13}$$

(c) the baryon conservation equation

$$n\left(1 - \frac{2m}{r}\right)^{\frac{1}{2}}\left\{\left(1 - \frac{2m}{r}\right)^{\frac{1}{2}}\frac{1}{r^2}\frac{\partial}{\partial r}(r^2 v^{(r)}) + \frac{1}{r\sin\theta}\right.$$
$$\left.\times\left[\frac{\partial}{\partial\theta}(\sin\theta\, v^{(\theta)}) + \frac{\partial v^{(\varphi)}}{\partial\varphi}\right]\right\} + \frac{Dn}{Dt} - \frac{n}{c^2(\rho + p/c^2)}$$
$$\times\left[\frac{Dp}{Dt} - \left(1 - \frac{v^2}{c^2}\right)\frac{\partial p}{\partial t}\right] = 0, \tag{14}$$

and

(d) the adiabatic equation

$$\frac{D}{Dt}(pn^{-\Gamma}) = 0, \tag{15}$$

wherein

$$v^2 = v^{(r)2} + v^{(\theta)2} + v^{(\phi)2}, \tag{16}$$

and

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \left(1 - \frac{2m}{r}\right)^{\frac{1}{2}}\left\{\left(1 - \frac{2m}{r}\right)^{\frac{1}{2}}v^{(r)}\frac{\partial}{\partial r} + \frac{v^{(\theta)}}{r}\frac{\partial}{\partial\theta} + \frac{v^{(\varphi)}}{r\sin\theta}\frac{\partial}{\partial\varphi}\right\}. \tag{17}$$

2.2 Steady state

While the above equations are very general, we impose the conditions of axi-symmetry and further that $v_0^{(r)} = 0$, $v_0^{(\theta)} = 0$, $v_0^{(\varphi)} \equiv v_0$ to consider the steady state. The equations governing the steady state are

$$\left(\rho_0 + \frac{p_0}{c^2}\right)\left[\frac{mc^2}{r^2} - \left(1 - \frac{2m}{r}\right)\frac{v_0^2}{r}\right] = -\left(1 - \frac{2m}{r}\right)\left(1 - \frac{v_0^2}{c^2}\right)\frac{\partial p_0}{\partial r}, \tag{18}$$

$$\left(\rho_0 + \frac{p_0}{c^2}\right)\cot\theta\, v_0^2 = \left(1 - \frac{v_0^2}{c^2}\right)\frac{\partial p_0}{\partial \theta}, \tag{19}$$

which connect three independent variables. One immediately finds that if $p_0 = 0$ then $\theta = \pi/2$ and $v_0^2 = (1 - 2m/r)^{-1} mc^2/r$ or that a pressureless disk collapses to $\theta = \pi/2$ plane executing Keplerian motion. Thus a non-zero pressure, necessarily implies a structured disk.

We first examine the situation in Newtonian gravitational field (by taking the limit $c \to \infty$). The steady state equations are relatively simpler in structure and their solutions have been obtained for a density distribution of the type $\rho_0 = \rho_c R^l$, as given by

$$\left(\frac{v_0}{c}\right)^2 = \frac{A}{R^{l-k}}\sin^k\theta, \tag{20}$$

$$\frac{p_0}{c^2} = \rho_c\left\{-\frac{R^{l-1}}{l-1} + \frac{A}{k}R^k\sin^k\theta + B\right\}, \qquad \text{when } l \neq 1, k \neq 0, \tag{21}$$

$$\frac{p_0}{c^2} = \rho_c\left\{-\frac{R^{l-1}}{l-1} + A\ln(R\sin\theta) + B\right\}, \qquad \text{when } l \neq 1, k = 0, \tag{22}$$

$$\frac{p_0}{c^2} = \rho_c\left\{-\ln R + \frac{A}{k}R^k\sin^k\theta + B\right\}, \qquad \text{when } l = 1, k \neq 0, \tag{23}$$

where l and k are constants and $R = r/m$. A and B are two dimensionless constants of integration, which can be determined by putting the boundary condition $p_0 = 0$ at the edges $R = a$ and $R = b$ of the disk on the equatorial plane of the black hole. The solutions obtained above are physically plausible provided $p_0 > 0$ within the interior of the disk and goes over to zero at its boundary. This leads to a constraint that $k < l - 1$ which implies that the azimuthal velocity v_0 should decrease with increasing R.

For the general relativistic case we obtained the solutions of steady state equations (18) and (19) by assuming ρ_0 to be constant and are given by

$$\left(\frac{v_0}{c}\right)^2 = \frac{A(1 - 2/R)}{R^2\sin^2\theta}, \tag{24}$$

$$\frac{p_0}{c^2} = \rho_0\left[B\left\{\left(1 - \frac{2}{R}\right)^{-1} - \frac{A}{R^2\sin^2\theta}\right\}^{\frac{1}{2}} - 1\right]. \tag{25}$$

The condition $p_0 > 0$ within the interior of the disk, now imposes a restriction on the inner edge. We find that a cannot be less than 4 and further

$$b > \frac{2a}{a-4}, \quad \text{if} \quad 4 < a < 6. \tag{26}$$

There is no restriction on the outer edge if $a \geqslant 6$. Figures 1 and 2 show the meridional section of the disk and figures 3 and 4 show the profiles of velocity and pressure.

3. Stability analysis

3.1 *Initial value equations and pulsation equations*

We consider the axisymmetric perturbations of the disk as discussed above and use the normal mode analysis restricting the perturbations to linear terms only. Defining the Lagrangian displacement ξ^α, $(\alpha = r, \theta)$ through

$$\delta v^{(\alpha)} = \frac{\partial \xi^\alpha}{\partial t}, \tag{27}$$

where δ denotes Eulerian perturbations and the time-dependence of all the perturbed variables to go as $\exp(i\sigma t)$ we obtain the following set of initial value equations (after integrating once with respect to time)

$$\left(\rho_0 + \frac{p_0}{c^2}\right)\delta v^{(\varphi)} = -S_2 \frac{v_0}{c^2}\delta p - \left(\rho_0 + \frac{p_0}{c^2}\right)\left\{\frac{\sqrt{S_1}}{r}\left(\frac{\partial v_0}{\partial \theta} + v_0 \cot\theta\right)\xi^\theta + \left(S_1 \frac{\partial v_0}{\partial r} + \frac{1}{r}\left[1 - \frac{3m}{r}\right]v_0\right)\xi^r\right\}. \tag{28}$$

$$\delta\rho = -\left(\rho_0 + \frac{p_0}{c^2}\right)\sqrt{S_1}\left[\frac{\sqrt{S_1}}{r^2}\frac{\partial}{\partial r}(r^2\xi^r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\,\xi^\theta)\right]$$

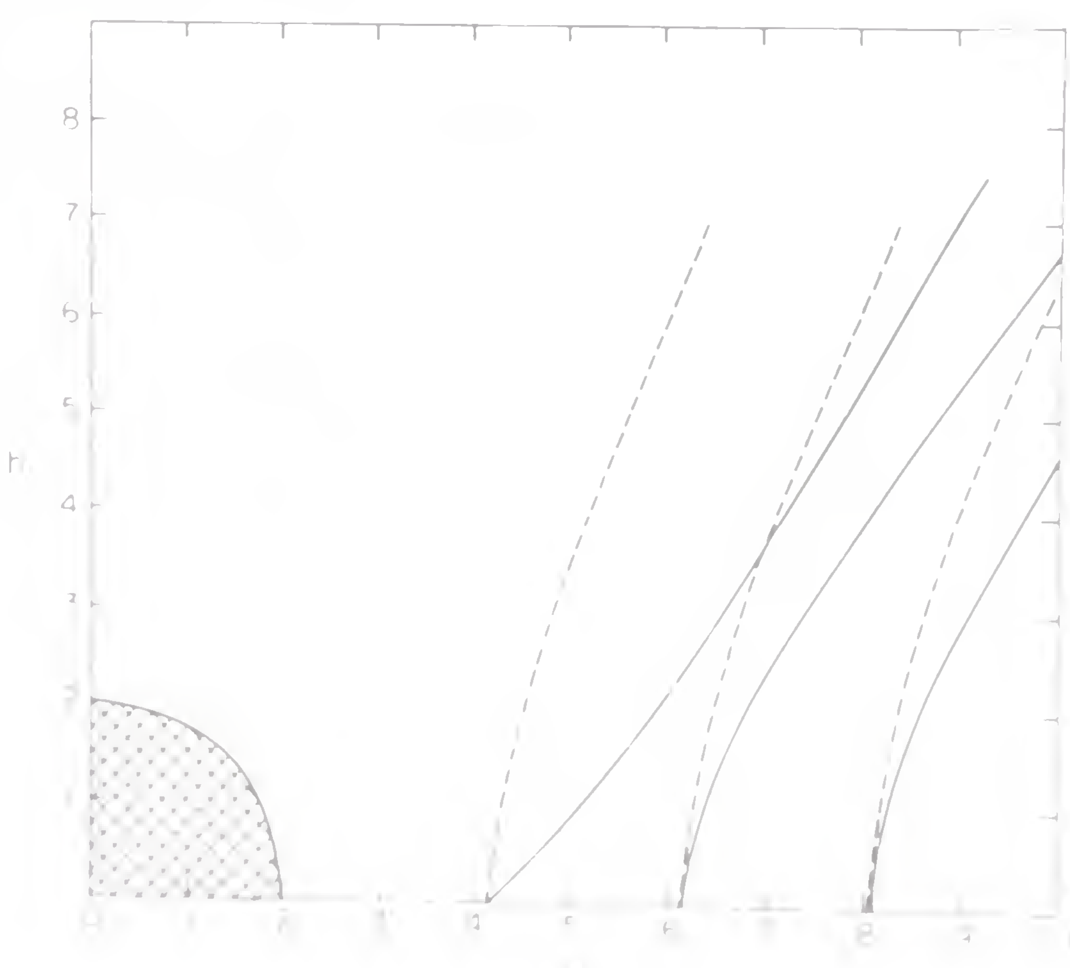

Figure 1. Inner portion of the meridional section of the disk for $b = 100$ and for various choices of a for general relativistic (solid curves) and Newtonian case (dashed curves).

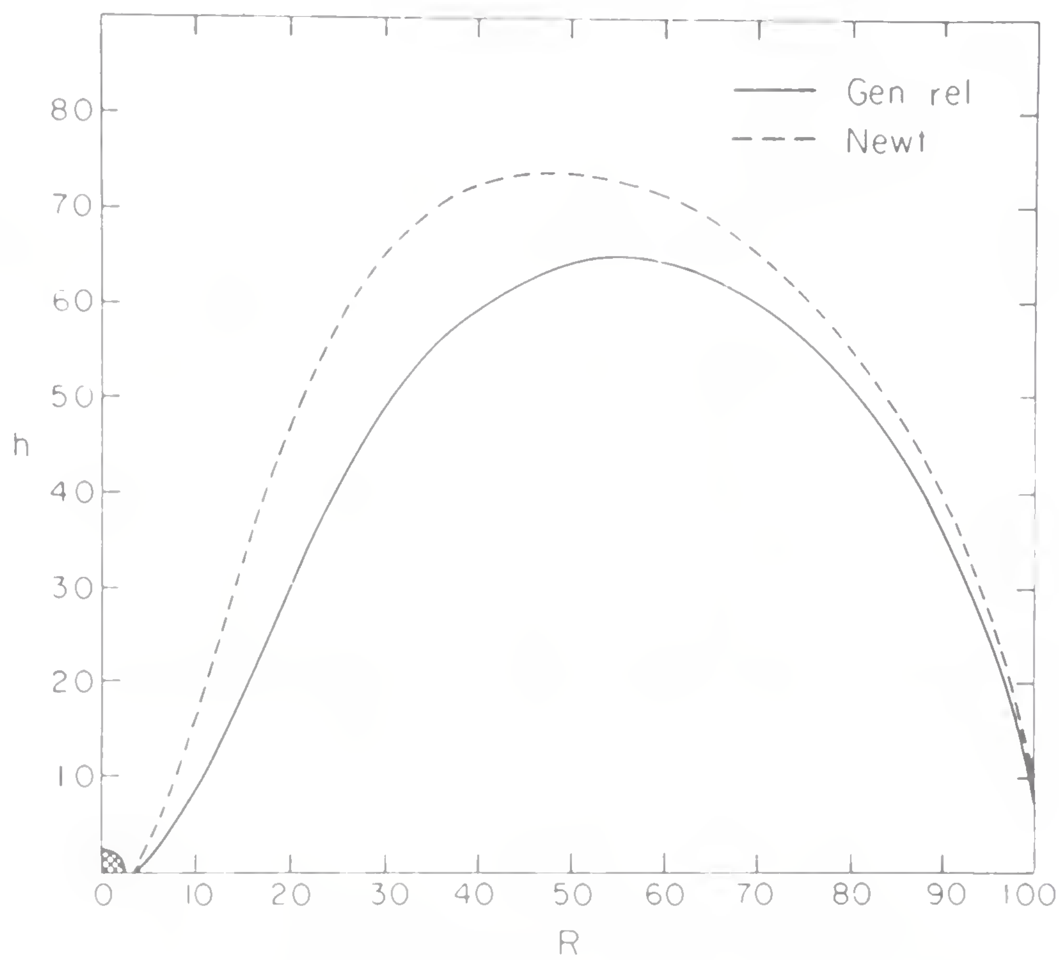

Figure 2. Meridional section of the disk in general relativistic (solid curve) and in Newtonian case (dashed curve) for $a = 4.1$, $b = 100$.

$$+\frac{v_0^2}{c^4}\delta p-\sqrt{S_1}\left\{\sqrt{S_1}\xi^r\frac{\partial}{\partial r}+\frac{\xi^\theta}{r}\frac{\partial}{\partial \theta}\right\}(\rho_0-p_0/c^2), \tag{29}$$

$$\delta n = -n_0\sqrt{S_1}\left[\frac{\sqrt{S_1}}{r^2}\frac{\partial}{\partial r}(r^2\xi^r)+\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\,\xi^\theta)\right]$$
$$-\sqrt{S_1}\left\{\sqrt{S_1}\xi^n\frac{\partial}{\partial r}+\frac{\xi^\theta}{r}\frac{\partial}{\partial\theta}\right\}n_0$$
$$+\frac{n_0}{c^2(\rho_0+p_0/c^2)}\left[\frac{v_0^2}{c^2}\delta p+\sqrt{S_1}\left\{\sqrt{S_1}\xi^r\frac{\partial}{\partial r}+\frac{\xi^\theta}{r}\frac{\partial}{\partial\theta}\right\}p_0\right], \tag{30}$$

and

$$\delta p = \frac{\gamma p_0}{n_0}\delta n-\sqrt{S_1}\left\{\sqrt{S_1}\xi^r\frac{\partial}{\partial r}+\frac{\xi^\theta}{r}\frac{\partial}{\partial\theta}\right\}p_0$$
$$+\frac{\gamma p_0}{n_0}\sqrt{S_1}\left\{\sqrt{S_1}\xi^r\frac{\partial}{\partial r}+\frac{\xi^\theta}{r}\frac{\partial}{\partial\theta}\right\}n_0, \tag{31}$$

and the following pair of pulsation equations

$$-\left(\rho_0+\frac{p_0}{c^2}\right)\sigma^2\xi^r=\left(\rho_0+\frac{p_0}{c^2}\right)\frac{2}{r}S_1v_0\delta v^{(\varphi)}-\left(\delta\rho+\frac{\delta p}{c^2}\right)\left[\frac{mc^2}{r^2}-\frac{S_1}{r^2}v_0^2\right]$$
$$-S_1S_2\frac{\partial}{\partial r}\delta p+2S_1\frac{v_0}{c^2}\frac{\partial p_0}{\partial r}\delta v^{(\varphi)}, \tag{32}$$

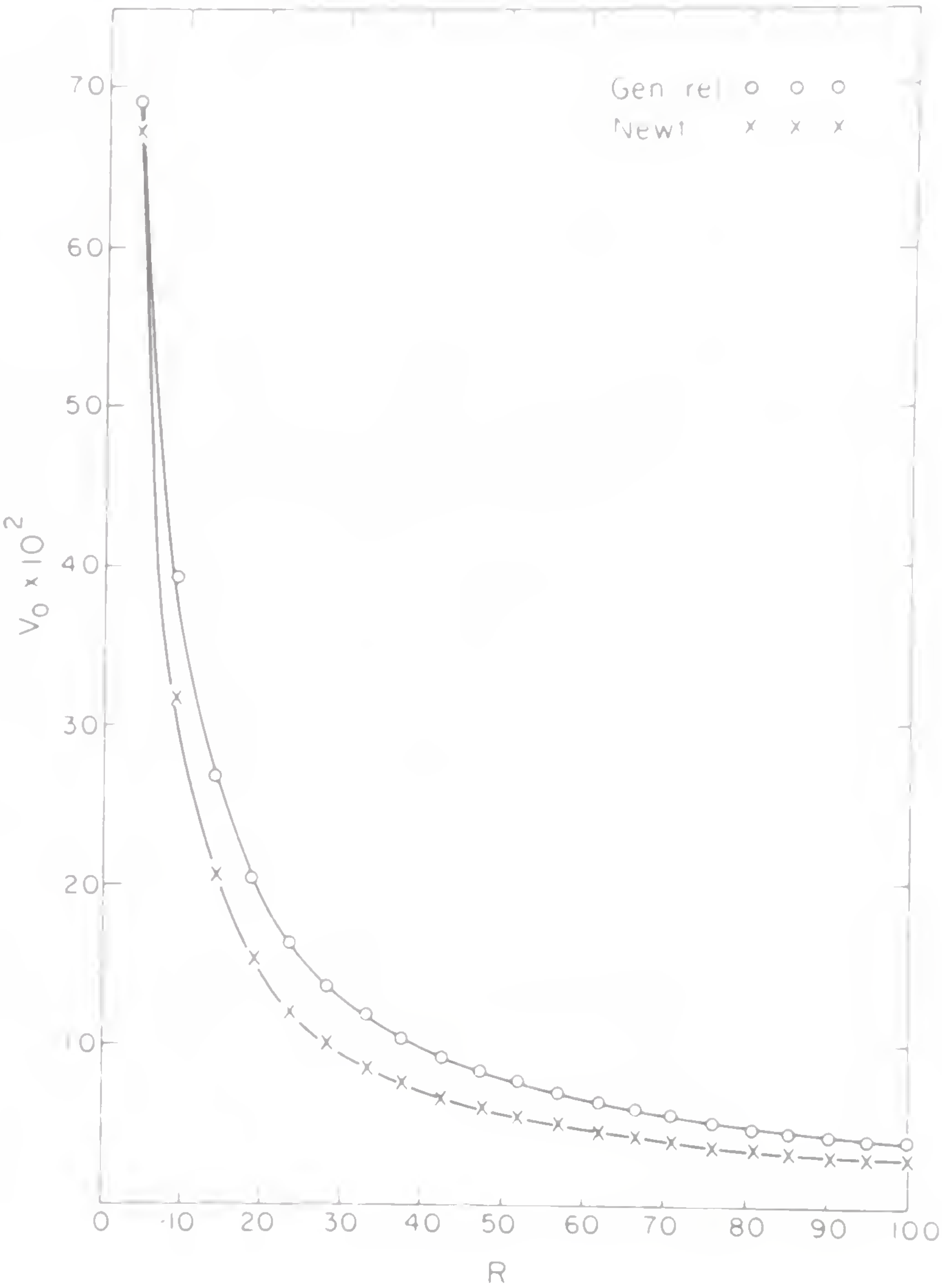

Figure 3. Profiles of velocity for relativistic (circles) and for Newtonian (crosses) disk along the equatorial plane for $a = 4.1$, $b = 100$.

and

$$-\left(\rho_0+\frac{p_0}{c^2}\right)\sigma^2\xi^\theta=\left(\rho_0+\frac{p_0}{c^2}\right)\frac{2}{r}\sqrt{S_1}\cot\theta v_0\delta v^{(\varphi)}+\left(\delta\rho+\frac{\delta p}{c^2}\right)\frac{v_0^2}{r}\sqrt{S_1}\cot\theta$$
$$-\sqrt{S_1S_2}\frac{1}{r}\frac{\partial}{\partial\theta}\delta p+2\sqrt{S_1}\frac{v_0}{c^2}\frac{\partial p_0}{\partial\theta}\frac{\delta v^{(\varphi)}}{r}. \tag{33}$$

wherein

$$S_1=\left(1-\frac{2m}{r}\right),\ S_2=\left(1-\frac{v_0^2}{c^2}\right), \tag{34}$$

and all the perturbed variables represent their spatial parts only. Equation (29) together with (30) yields

$$\frac{\Delta\rho}{\rho_0+p_0/c^2}=\frac{\Delta n}{n_0} \tag{35}$$

while (31) can be rewritten as

$$\frac{\Delta p}{p_0}=\gamma\frac{\Delta n}{n_0}. \tag{36}$$

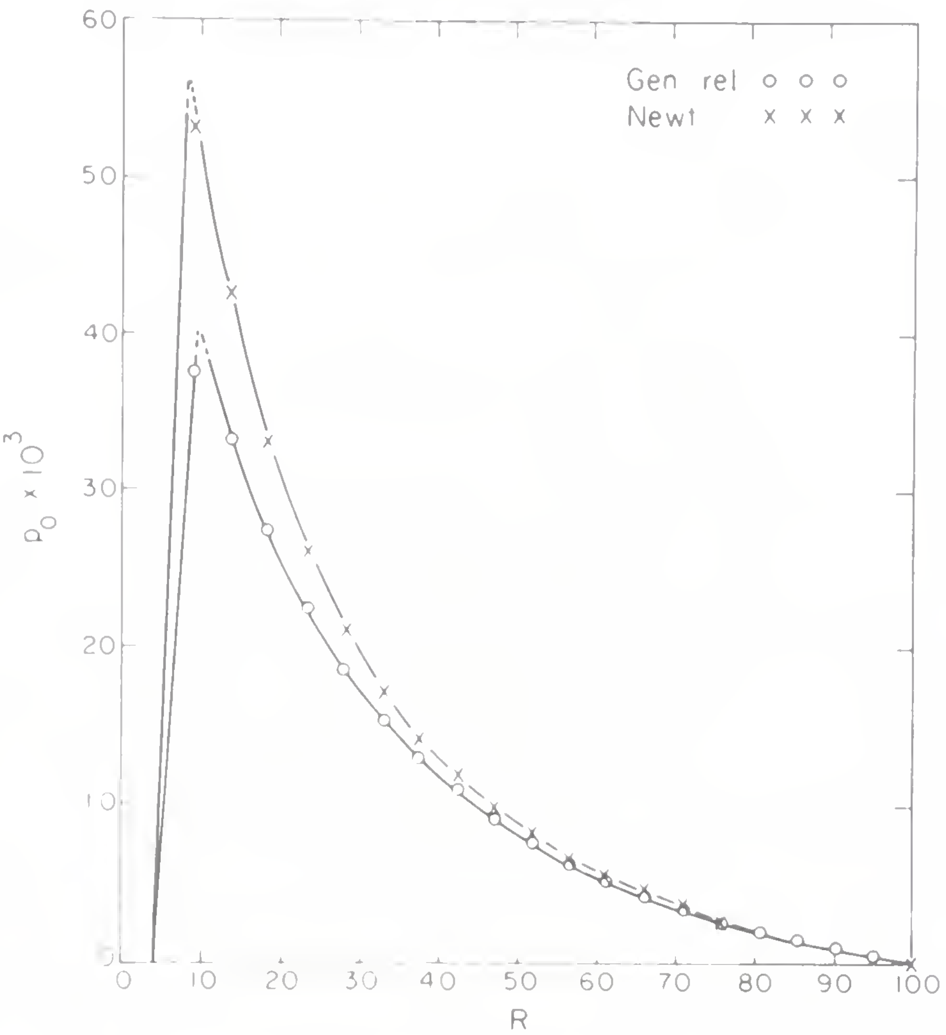

Figure 4. Profiles of pressure for relativistic (circles) and for Newtonian (crosses) disks along the equatorial plane for $a = 4.1$, $b = 100$.

in terms of Lagrangian perturbations denoted by Δ. Our problem is to solve (32) and (33) as eigenvalue equations, consistently with the initial value equations and appropriate boundary conditions. At the edges of the disk we need the boundary condition $\Delta p = 0$, which is satisfied by requiring ξ^α and its derivatives to remain finite everywhere.

From (29) to (31) we obtain

$$\delta p\left[1-\frac{\gamma p_0/c^2}{\rho_0+p_0/c^2}\frac{v_0{}^2}{c^2}\right]=-\left[1-\frac{\gamma p_0/c^2}{\rho_0+p_0/c^2}\right]\left(S_1\xi^r\frac{\partial p_0}{\partial r}+\sqrt{S_1}\,\frac{\xi^\theta}{r}\frac{\partial p_0}{\partial\theta}\right)$$
$$-\gamma p_0\left[\frac{S_1}{r^2}\frac{\partial}{\partial r}(r^2\xi^r)+\frac{\sqrt{S_1}}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\xi^\theta)\right], \qquad (37)$$

$$\delta\rho\left[1-\frac{\gamma p_0/c^2}{\rho_0+p_0/c^2}\frac{v_0^2}{c^2}\right]=-\left(\rho_0+\frac{p_0}{c^2}\right)\left[\frac{S_1}{r^2}\frac{\partial}{\partial r}(r^2\xi^r)+\frac{\sqrt{S_1}}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\xi^\theta)\right]$$
$$-\left(S_1\xi^r\frac{\partial\rho_0}{\partial r}+\sqrt{S_1}\,\frac{\xi^\theta}{r}\frac{\partial\rho_0}{\partial\theta}\right)\left[1-\frac{\gamma p_0/c^2}{\rho_0+p_0/c^2}\frac{v_0^2}{c^2}\right]$$
$$+\frac{S_2}{c^2}\left(S_1\xi^r\frac{\partial p_0}{\partial r}+\sqrt{S_1}\frac{\xi^\theta}{r}\frac{\partial p_0}{\partial\theta}\right) \qquad (38)$$

which along with (28) form alternative expressions for the initial value equations.

3.2 *Equation for σ^2*

Following the procedure of Chandrasekhar and Friedman (1972a, b), we multiply the pulsation equation (32) by $\bar{\xi}^r$ and (33) by $\bar{\xi}^\theta$, add them and integrate with respect to r and θ over the entire region of the disk to obtain an expression of the form

$$\sigma^2 \int\int\left(\rho_0 + \frac{p_0}{c^2}\right)\{\xi^r\bar{\xi}^r + \xi^\theta\bar{\xi}^\theta\}\, r^2 \sin\theta \,\mathrm{d}r\,\mathrm{d}\theta$$

$$= \int\int (I)\, r^2 \sin\theta\,\mathrm{d}r\,\mathrm{d}\theta. \qquad (39)$$

Here $\bar{\xi}^r$ and $\bar{\xi}^\theta$ are the 'trial functions' which satisfy the same boundary conditions as required by the true eigen-functions ξ^r and ξ^θ, but otherwise completely arbitrary. We also define barred variations $\Delta\bar{\rho}$, $\Delta\bar{p}$ etc. as obtained from the initial value equations when ξ^α is replaced by $\bar{\xi}^\alpha$. We find that the left side of (39) is manifestly symmetric in barred and unbarred variables. By performing several integrations by parts and using the steady state equations, initial value equations and the condition that $p_0 = 0$ at the boundary, we bring the right side of (39) in a form which is manifestly symmetric in barred and unbarred variables. Identifying the barred variables with the unbarred ones we get an equation for which for the case of Newtonian gravitational field is given by

$$\frac{m^2\sigma^2}{c^2}\int\int \rho_0 R^2 \sin\theta(\hat{\xi}^{r2} + \hat{\xi}^{\theta 2})\,\mathrm{d}R\,\mathrm{d}\theta$$

$$= \int\int\Bigg[2\rho_0\frac{v_0}{c^2}R\sin\theta\left(\frac{\partial v_0}{\partial R} + \frac{v_0}{R}\right)\hat{\xi}^{r2} - \frac{\partial p_0}{\partial R}\sin\theta\,\hat{\xi}^{r2}$$

$$- R^4\sin\theta\frac{\hat{\xi}^{r2}}{c^2}\frac{\partial}{\partial R}\left(\frac{1}{R^2}\frac{\partial p_0}{\partial R}\right) + \left(\frac{v_0}{c}\right)^2 R\sin\theta\,\hat{\xi}^{r2}\frac{\partial \rho_0}{\partial R}$$

$$+ 2\rho_0\frac{v_0}{c^2}\cos\theta\left(\frac{\partial v_0}{\partial\theta} + \cos\theta v_0\right)\hat{\xi}^{\theta 2}$$

$$- \sin^2\theta\frac{\hat{\xi}^{\theta 2}}{c^2}\frac{\partial}{\partial\theta}\left(\frac{1}{\sin\theta}\frac{\partial p_0}{\partial\theta}\right) + \hat{\xi}^{\theta 2}\left(\frac{v_0}{c}\right)^2\cos\theta\frac{\partial\rho_0}{\partial\theta}$$

$$+ 4\rho_0\left(\frac{v_0}{c}\right)^2\cos\theta\,\hat{\xi}^r\hat{\xi}^\theta - \frac{2p_0}{c^2}\frac{\partial}{\partial R}(R\hat{\xi}^r)\frac{\partial}{\partial\theta}(\sin\theta\hat{\xi}^\theta)$$

$$+ \frac{2p_0}{c^2}\frac{\partial}{\partial\theta}(R\hat{\xi}^r)\frac{\partial}{\partial R}(\sin\theta\,\hat{\xi}^\theta) + \frac{2}{c^2}\frac{\partial p_0}{\partial\theta}\sin\theta\,\hat{\xi}^r\hat{\xi}^\theta\Bigg]\mathrm{d}R\,\mathrm{d}\theta$$

$$+ \gamma\int\int\Bigg[\frac{p_0}{R^2c^2}\sin\theta\left\{\frac{\partial}{\partial R}(R^2\hat{\xi}^r)\right\}^2 + \frac{p_0}{\sin\theta c^2}\left\{\frac{\partial}{\partial\theta}(\sin\theta\,\hat{\xi}^\theta)\right\}^2$$

$$+ \frac{2p_0}{Rc^2}\frac{\partial}{\partial R}(R^2\hat{\xi}^r)\frac{\partial}{\partial\theta}(\sin\theta\,\hat{\xi}^\theta)\Bigg]\mathrm{d}R\,\mathrm{d}\theta \qquad (40)$$

where $\hat{\xi}^\alpha = \xi^\alpha/m$

The right side of (40) has been written as the coefficients of various powers of γ. To

obtain σ^2-equation for the general relativistic case, in a similar form, we limit ourselves to the situations where

$$\frac{\gamma p_0/c^2}{(\rho_0 + p_0/c^2)} \frac{v_0^2}{c^2} \ll 1$$

and also we impose a slightly more restrictive boundary condition that $\delta p = 0$ at the boundary. This boundary condition is fulfilled for any ξ^α which is zero at the boundary. We then obtain the following expression for σ^2 (using (24) and that $\rho_0 =$ constant)

$$\begin{aligned}
\frac{m^2\sigma^2}{c^2} & \iint \frac{1}{S_2}\left(\rho_0 + \frac{p_0}{c^2}\right)(\hat{\xi}^{r2} + \hat{\xi}^{\theta 2}) R^2 \sin\theta \mathrm{d}R\,\mathrm{d}\theta \\
&= \iint \Bigg[\frac{v_0^2/c^2}{\rho_0 + p_0/c^2} T_2^2 - 2S_1 \hat{\xi}^r \frac{\partial T_2}{\partial R} - \frac{2\sqrt{S_1}}{R} \hat{\xi}^\theta \frac{\partial T_2}{\partial R} \\
&\quad - \frac{4}{R^2} \hat{\xi}^r T_2 \Bigg] R^2 \sin\theta \,\mathrm{d}R\, d\theta \\
&\quad + \gamma \iint \Bigg[-2S_1 \hat{\xi}^r \frac{\partial}{\partial R}\left(T_1 \frac{p_0}{c^2}\right) + 2S_1 \hat{\xi}^r \frac{\partial}{\partial R}(S_2 S_3 T_2) \\
&\quad - \frac{2\sqrt{S_1}}{R} \hat{\xi}^\theta \frac{\partial}{\partial\theta}\left(T_1 \frac{p_0}{c^2}\right) + \frac{2\sqrt{S_1}}{R} \hat{\xi}^\theta \frac{\partial}{\partial\theta}(S_2 S_3 T_2) \\
&\quad - \frac{4}{R^2} \hat{\xi}^r \left(\frac{p_0}{c^2} T_1 - S_2 S_3 T_2\right) - 2S_2 S_3 \{T_1 T_2 - S_2 T_2^2/(\rho_0 + p_0/c^2)\} \\
&\quad - S_3 \left\{ \left(\rho_0 + \frac{p_0}{c^2}\right) T_1^2 + \frac{S_2^2 T_2^2}{(\rho_0 + p_0/c^2)} - 2S_2 T_1 T_2 \right\} \Bigg] R^2 \sin\theta \,\mathrm{d}R\,\mathrm{d}\theta \\
&\quad + \gamma^2 \iint \Bigg[-2S_1 \hat{\xi}^r \frac{\partial}{\partial R}\left(S_5 T_1 \frac{p_0}{c^2}\right) + 2S_1 \hat{\xi}^r \frac{\partial}{\partial R}(S_4^2 T_2) \\
&\quad - 2\frac{\sqrt{S_1}\hat{\xi}^\theta}{R} \frac{\partial}{\partial\theta}\left(S_5 T_1 \frac{p_0}{c^2}\right) + \frac{2\sqrt{S_1}\hat{\xi}^\theta}{R} \frac{\partial}{\partial\theta}(S_4^2 T_2) \\
&\quad + \frac{4}{R^2} \hat{\xi}^r \left(S_4^2 T_2 - \frac{p_0}{c^2} T_1 S_5\right) - 2S_1 S_3 \left\{ S_5 T_1 T_2 - \frac{S_2 S_5}{(\rho_0 + p_0/c^2)} T_2^2 \right\} \\
&\quad + S_3 S_5 \left\{ \left(\rho_0 + \frac{p_0}{c^2}\right) T_1^2 + \frac{S_2^2 T_2^2}{(\rho_0 + p_0/c^2)} - 2S_2 T_1 T_2 \right\} \\
&\quad - S_3 \left\{ 2\left(\rho_0 + \frac{p_0}{c^2}\right) S_5 T_1^2 + \frac{2S_2^2 S_5 T_2^2}{(\rho_0 + p_0/c^2)} - 4S_5 S_2 T_1 T_2 \right\} \Bigg] R^2 \sin\theta \,\mathrm{d}R\,\mathrm{d}\theta \\
&\quad + \gamma^3 \iint \Bigg[-S_3 \left\{ \left(\rho_0 + \frac{p_0}{c^2}\right) S_5^2 T_1^2 + \frac{S_5^2 S_2^2 T_2^2}{(\rho_0 + p_0/c^2)} - 2S_5 S_2 T_1 T_2 \right\} \\
&\quad + S_3 S_5 \left\{ 2\left(\rho_0 + \frac{p_0}{c^2}\right) S_5 T_1^2 + \frac{2S_2^2 S_5 T_2^2}{(\rho_0 + p_0/c^2)} - 4S_5 S_2 T_1 T_2 \right\} \Bigg] R^2 \sin\theta \,\mathrm{d}R\,\mathrm{d}\theta
\end{aligned}$$

$$+\gamma^4 \iint \left[S_3 S_5 \left\{ \left(\rho_0 + \frac{p_0}{c^2}\right) S_5^2 T_1^2 + \frac{S_5^2 S_2^2 T_2^2}{(\rho_0 + p_0/c^2)} - 2 S_2 S_5^2 T_1 T_2 \right\} \right] R^2 \sin\theta \, \mathrm{d}R \, \mathrm{d}\theta \tag{41}$$

wherein

$$S_3 = \frac{p_0/c^2}{\rho_0 + p_0/c^2}, \; S_4 = S_3 v_0/c, \; S_5 = S_4 v_0/c$$

$$T_1 = \frac{S_1}{R^2} \frac{\partial}{\partial R} (R^2 \hat{\xi}^r) + \frac{\sqrt{S_1}}{R \sin\theta} \frac{\partial}{\partial\theta} (\sin\theta \, \hat{\xi}^\theta)$$

$$T_2 = S_1 \hat{\xi}^r \frac{\partial}{\partial R}\left(\frac{p_0}{c^2}\right) + \sqrt{S_1} \frac{\hat{\xi}^\theta}{R} \frac{\partial}{\partial\theta}\left(\frac{p_0}{c^2}\right). \tag{42}$$

The above expressions for σ^2 imply a variational principle in the following sense: if we evaluate (40) or (41) by two trial displacements $\hat{\xi}^\alpha$ and $\hat{\xi}^\alpha + \delta\hat{\xi}^\alpha$ such that the variation in σ^2 is $\delta\sigma^2$ and demand that the variation $\delta\sigma^2$ in σ^2 is zero, then it amounts to solving the original set of eigen-value equations (cf: Chandrasekhar and Friedman 1972b, Chakraborty and Prasanna 1982).

3.3 *Neutral mode*

Using the equations for σ^2 as obtained above, we calculate the critical value γ_c, of the adiabatic index for $\sigma^2 = 0$. The procedure is straightforward. We choose trial functions for ξ^r and ξ^θ which contain several adjustable parameters $\alpha, \beta \ldots$ etc. and calculate the integrals numerically which appear on the right side of σ^2-equation, as being the coefficients of various powers of γ. If $Y_1, Y_2 \ldots$ etc. are the integrals as obtained by numerical integration, we have

$$\sigma^2 \iint \ldots \mathrm{d}R \, \mathrm{d}\theta = Y_1 + \gamma_c Y_2 + \gamma_c^2 Y_3 + \ldots \tag{43}$$

Each integral $Y_1, Y_2 \ldots$ etc. contains the parameters $\alpha, \beta \ldots$ etc. which can be determined by extremising (43) with respect to these parameters. With these values of the parameters, we determine γ_c from (43) by putting $\sigma^2 = 0$.

For the case of Newtonian gravitational field, we choose two kinds of trial functions (i) with fixed boundary *i.e.* the Lagrangian displacement ξ^α vanishing at the boundary, and (ii) with non-stationary boundary. (i) We choose a function q which vanishes at the boundary. This function can easily be chosen from the form of p_0. Thus for the Newtonian case, corresponding to the three different types of solutions (47) to (49) we choose

$$q = \sin^k\theta - \frac{k}{AR^k}\left(\frac{R^{l-1}}{l-1} - B\right), \quad l \neq 1, k \neq 0,$$

$$q = \sin\theta - \exp\left\{-\frac{R^{l-1}}{A(l-1)} - \frac{B}{A} - \ln R\right\}, \quad l \neq 1, k = 0,$$

$$q = \sin^k\theta - \frac{k}{AR^k}(\ln R - B), \quad l = 1, k \neq 0 \tag{44}$$

We choose $\hat{\xi}^\alpha$ as

$$\hat{\xi}^r = q + \alpha q^2, \ \hat{\xi}^\theta = q + \beta q^2 \tag{45}$$

and determine γ_c by extremizing σ^2, as given by (40), with respect to α and β as indicated above.

(ii) In case of non-stationary boundary we first consider the case of radial perturbation with $\xi^\theta = 0$. The equations governing such radial perturbations are obtained from (27), (37), (38), (33) and (34) by putting $\xi^\theta = 0$ and taking the limit $c \to \infty$ and are as follows:

$$\delta v^\varphi = -\left(\frac{\partial v_0}{\partial r} + \frac{v_0}{r}\right)\xi^r, \tag{46}$$

$$\delta\rho = -\frac{\rho_0}{r^2}\frac{\partial}{\partial r}(r^2\xi^r) - \xi^r\frac{\partial\rho_0}{\partial r}, \tag{47}$$

$$\delta p = -\frac{\gamma p_0}{r^2}\frac{\partial}{\partial r}(r^2\xi^r) - \xi^r\frac{\partial p_0}{\partial r}, \tag{48}$$

$$-\rho_0\sigma^2\xi^r = \frac{2\rho_0 v_0}{r}\delta v^\varphi - \left(\frac{MG}{r^2} - \frac{v_0^2}{r}\right)\delta\rho - \frac{\partial}{\partial r}\delta p, \tag{49}$$

$$\frac{2\rho_0 v_0}{r}\cot\theta\,\delta v\varphi + \frac{v_0^2}{r}\cot\theta\,\delta\rho - \frac{1}{r}\frac{\partial}{\partial\theta}\delta p = 0. \tag{50}$$

Using the initial value equations (46) to (48) in (50) and assuming $\xi^r = \xi^r(r)$, we obtain a differential equation

$$(1-\gamma)\frac{\mathrm{d}}{\mathrm{d}r}(r^2\xi^r) + 2r\xi^r = 0, \tag{51}$$

for ξ^r, whose solution is given by

$$\xi^r = \eta r^{\left(\frac{4-2\gamma}{\gamma-1}\right)} \tag{52}$$

where η is a constant. Using this solution for ξ^r in (49) we get

$$\rho_0\frac{m^2\sigma^2}{c^2}\xi^r = \rho_c\left\{\frac{2\gamma(3\gamma-5)}{(\gamma-1)^2}B + A\sin^k\theta\,\frac{(3\gamma-5)(k\gamma-k+2\gamma)}{k(\gamma-1)^2}R^k + \left[\frac{4-2\gamma}{\gamma-1} - \frac{2\gamma}{(\gamma-1)(l-1)}\left(\frac{3\gamma-5}{\gamma-1}\right)\right]R^{l-1}\right\} \times m^2\eta r^{\left(\frac{6-4\gamma}{\gamma-1}\right)},\ l\neq 1,\ k\neq 0 \tag{53}$$

$$\rho_0\frac{m^2\sigma^2}{c^2}\xi^r = \rho_c\left\{\frac{2\gamma(3\gamma-5)}{(\gamma-1)^2}B + 2A\ln(R\sin\theta)\frac{(3\gamma-5)}{(\gamma-1)^2} + A\frac{(3\gamma-5)}{(\gamma-1)} + \left[\frac{4-2\gamma}{\gamma-1} - \frac{2\gamma}{(\gamma-1)(l-1)}\left(\frac{3\gamma-5}{\gamma-1}\right)\right] \times R^{l-1}\right\} m^2\eta r^{\left(\frac{6-4\gamma}{\gamma-1}\right)},\ l=1,\ k=0 \tag{54}$$

$$\rho_0 \frac{m^2\sigma^2}{c^2}\,\xi^r = \rho_c \left\{ \frac{2\gamma(3\gamma-5)}{(\gamma-1)^2} B + A \sin^k \theta \frac{(3\gamma-5)(k\gamma-k+2\gamma)}{k(\gamma-1)^2} R^k \right.$$
$$\left. + \left(\frac{4-2\gamma}{\gamma-1}\right) - (\ln R)\frac{2\gamma}{\gamma-1}\left(\frac{3\gamma-5}{\gamma-1}\right)\right\}$$
$$\times\, m^2 \eta r^{\left(\frac{6}{\gamma}-\frac{4}{1}\right)}, \; l = 1,\; k \neq 0. \tag{55}$$

For the special case of ordinary gas with $\gamma = 5/3$ the above equations for σ^2 reduce to a very simple form

$$\frac{m^2\sigma^2}{c^2} = \frac{1}{R^3}, \tag{56}$$

showing that the disks are stable with 'local' frequency being proportional to $R^{-3/2}$ irrespective of the other parameters like l, k, a and b. The function ξ^r takes the form ηr for $\gamma = 5/3$ which is exactly the form as used by Bisnovatyi and Blinnikov (1972) for analysing the stability of thin gas disk against expansion and contraction. It is interesting to note that the frequency obtained above is also the same for the radial oscillation of a pressureless disk confined to $\theta = \pi/2$ plane and on Keplerian motion (non-relativistic) with $v_0 = (MG/r)^{1/2}$, as may be seen from (46) to (49) with $\delta p = 0$.

To consider the axisymmetric perturbation with non-stationary boundary we choose

$$\xi^r = R + \alpha q,\; \xi^\theta = R + \beta q \tag{57}$$

and calculate critical value γ_c of the adiabatic index for $\sigma^2 = 0$, as in the previous case. Table 1 shows the values of critical γ (denoted by γ_{c_1} and γ_{c_2} respectively for the cases of fixed and non-stationary boundaries) for various values of l and k and for $a = 4$, $b = 100$.

As we have used the boundary condition $\delta p = 0$ to obtain (41) for the general relativistic case, we limit ourselves to the perturbations with fixed boundary. The function q for this case is chosen as

$$q = \frac{1}{\sin^2\theta} - \frac{R^2\left[B^2(1-2/R)^{-1}-1\right]}{AB^2}. \tag{58}$$

Table 1 Critical values of γ for neutral stability for Newtonian disk with $a = 4{\cdot}0$, $b = 100$

l	k	γ_{c_1}	γ_{c_2}
−2	−4	1·05	1·10
−2	−5	1·06	1·17
−2	−6	1·02	1·11
−1	−3	1·00	1·09
−1	−4	1·04	1·17
−1	−5	1·00	1·11
0	−2	0·89	1·04
0	−3	0·99	1·16
0	−4	0·97	1·15
1	−1	0·62	0·93
1	−2	0·88	1·14
1	−3	0·92	1·15

and choosing ξ^{α} as in (45), we determine γ_c for the neutral stability as in the previous cases. Table 2 indicates the values of γ_c for various values of the parameters a and b. The table also indicates the values of γ_c for the corresponding Newtonian case.

It is interesting to consider the radial oscillations of a pressureless disk, which is governed by

$$\delta v^{(\varphi)} = -\left[\left(1-\frac{2m}{r}\right)\frac{\partial v_0}{\partial r} + \frac{1}{r}\left(1-\frac{3m}{r}\right)v_0\right]\xi^r \tag{59}$$

$$-\sigma^2\xi^r = \frac{2}{r}\left(1-\frac{2m}{r}\right)v_0\,\delta v(\varphi), \tag{60}$$

as may be obtained from (28) and (32) by putting $\xi^\theta = 0$, $p_0 = 0$, $\delta p = 0$. Combining these, we get

$$\sigma^2 = \frac{mc^2}{r^4}(r-6m) \tag{61}$$

which shows that the disks are stable for $r > 6m$.

4. Discussion

Profiles of the steady state parameters—velocity and pressure, as a function of radial distance along the equatorial plane of the black hole is presented in figures 3 and 4 while figures 1 and 2 show the meridional section of such disk (solid line is for general relativistic case while dotted line is for Newtonian disk), when density is constant. For the Newtonian disk we have presented the case referring to $l = 0$, $k = -2$, to which the general relativistic steady state solutions reduce to, when the limit $c \to \infty$ is taken. From figure 1 and 2 and also from the similar plots for other values of l, k, a and b (for Newtonian disk), we find that the disks with non-zero pressure and thick, occupying a considerably larger volume than the central black hole. For the same value of a and b, we find that the Newtonian disk occupies more volume than the relativistic one. It seems that the relativistic disks show the formation of cusp at the inner edge specially when it is

Table 2 Critical values of γ for general relativistic and Newtonian disk.

a	b	γ_c Gen. Rel.	γ_c Newt.
8.1	100	0·7488	0·7107
7·1	100	0·7758	0·7420
6·1	100	0·8007	0·7732
5.1	100	0·8214	0·8043
4·1	100	0·8320	0·8356
4·1	140	0·8850	0·8726
4·1	180	0·9155	0·8936
4·05	180	0·9155	0·8946

near $4m$. For the Newtonian disk, pressure at any point is higher while the velocity at any point is lower at the equatorial plane, than for the relativistic disk.

For the Newtonian disk we find the constraint that $k < l-1$ which means that the velocity should decrease with increasing R. This is quite reasonable from the point of view of balance of various forces. At the inner edge a the gravitational force is stronger and besides the pressure gradient forces are directed towards the central star. Hence the centrifugal force and therefore the velocity v_0 must be higher. Just the opposite happens at the outer edge b.

For the relativistic disk we find the constraint that the inner edge cannot lie inside $4m$. Further if $4 < a < 6$, $b > 2a/(a-4)$. For $a \geqslant 6$, any $b > a$ gives rise to physically plausible disks. No such restriction appears in the Newtonian formulations, indicating a pure general relativistic origin of the present constraint. This is consistent with the general proof given by Kozlowsky *et al* (1978), that $R_m \geqslant R_{mb}$.

As regards the onset of instability, as discussed earlier we calculate the critical adiabatic index γ_c for $\sigma^2 = 0$ for different values of the disk parameters. Tables 1 and 2 indicate that $\gamma_c < 4/3$ for all the cases that we considered indicating stable configurations, for the axisymmetric perturbations both for fixed and non-stationary boundaries (for Newtonian disk, Table 1) and for fixed boundary (for general relativistic disk, Table 2). In calculating γ_c, for the general relativistic disk, we have used the approximation $(v_0^2/c^2)(p_0/c^2)/(\rho_0 + p_0/c^2) \ll 1$ which is quite justified from the values of v_0/c and p_0/c^2 as we obtained. There is a qualitative agreement between the γ_c calculated for relativistic and for the corresponding Newtonian disks (table 2). In these calculations, although the inner and outer radii a and b are the same, the regions occupied by the disks in the two cases are not the same. In general, Newtonian disks are thicker (figures 1 and 2). We find from the numbers that γ_c depends upon the size of the disk. In the calculations of γ_c, the effects due to general relativistic corrections and that due to the difference in sizes have contributed simultaneously and therefore the agreement between the general relativistic γ_c and the Newtonian γ_c is no better than a qualitative one. It does not seem to be possible to separate the contributions from different effects in the present formulation.

For the case of radial perturbations we have found that an ordinary perfect fluid ($\gamma = 5/3$) disk is stable with the local frequency $(MG/r^3)^{1/2}$ in the Newtonian formulation (Kato and Fukue 1980, and Cox 1981, have also considered local and quasi-radial oscillations of a thin gaseous disk in Schwarzschild background, when $p_0 \ll \rho_0 c^2$).

For a pressureless thin disk collapsing to $\theta = \pi/2$ plane we found that the disk is stable under radial perturbation if the inner edge is beyond $6m$ (general relativistic case), with local frequency $[mc^2(r-6m)/r^4]^{1/2}$. Now since a pressureless fluid is essentially an aggregate of non-interacting particles, the above result can be regarded as an alternative derivation of a well-known result that the last stable circular orbit for Schwarzschild geometry is at $6m$. A pressureless thin disk in Newtonian formulation is stable under radial oscillations, for all values of r, with local frequency $(MG/r^3)^{1/2}$.

The general conclusion that the perfect fluid thick disk is constant density and rotating around a stationary black hole, are generally stable, may have important significance in the study of the models of accretion disks for the high energy sources. However, it is very important to note that the complete discussion of the disk dynamics would require the inclusion of self-consistent electromagnetic fields as well as the dissipative forces like viscosity and radiation pressure.

References

Abramowicz M, Calvani M and Nobili L 1980 *Astrophys. J.* **242** 772
Bisnovayi-Kogan G S and Blinnikov I 1972 *Astrophys. and Space Sci.* **19** 119
Chakraborty D K and Prasanna A R 1981 *J. Astrophys. Astron.* **2** 421
Chakraborty D K and Prasanna A R 1982 *J. Astrophys. Astron.* **3** 193
Cox J P 1981 *Astrophys. J.* **247** 1070
Chandrasekhar S and Friedman J L 1972a *Astrophys. J.* **175** 379
Chandrasekhar S and Friedman J L 1972b *Astrophys. J.* **176** 745
Fishbone L G and Moncrief V 1976 *Astrophys. J.* **207** 962
Jaroszynski M, Abramowicz M and Paczynski B 1980 *Acta. Astron.* **30** 1
Kato S and Fukue J 1980 *Publ. Astron. Soc. Jap.* **32** 377
Kozlowski M, Jaroszynski M and Abramowicz M 1978 *Astron. Astrophys.* **63** 209
Lightman A P 1974a *Astrophys. J.* **194** 419
Lightman A P 1974b *Astrophys. J.* **194** 429
Novikov I D and Thorne K S 1973 *Black holes* (eds.) B Dewitt and C Dewitt (New York: Gordon and Breach)
Paczynski B and Wiita P J 1980 *Astron. and Astrophys.* **88** 23
Prasanna A R and Chakraborty D K 1981 *J. Astrophys. Astron.* **2** 1
Pringle J E and Rees M J 1972 *Astron. and Astrophys.* **21** 1
Shakura N I and Sunyaev R A 1973 *Astron. Astrophys.* **24** 337
Shakura N I and Sunyaev R A 1976 *Mon. Not. R. Astr. Soc.* **175** 613
Wiita P J and TIFR group 1980 Preprint

Discussion

A. R. Prasanna: I think it is significant to point out the possibility of cusp formation in the case of thick disks.

D. K. Chakraborty: As the detailed analysis shows for the structure of the steady state disk, in the Newtonian formulation the inner edge appears smooth, whereas with the General relativistic formulation, the inner edge shows the development of a cusp. This cusp could be the mouth of the funnel that one talks about for feeding material to the black hole from the disk.

§ IV. CLASSICAL COSMOLOGY

INTRODUCTION

Whereas use of Einstein's theory of general relativity (GR) to understand astrophysical phenomena dates back to early sixties, the application of the theory to understand the overall picture of the universe began almost with the theory itself. What are now termed as standard cosmological models were first discussed by Friedman way back in early twenties. As a matter of fact, before the epoch-making astronomical discoveries of quasars, pulsars and the microwave background radiation, the only area in which GR showed any potentiality of breaking new ground was cosmology. Several alternate (non-Einsteinian) theories were also proposed during this period to explain the rather meagre observational data on the universe as a whole. However with the development of new branches of astronomy—radio astronomy, x-ray and γ-ray astronomy—the observational data on the universe as a whole increased considerably and at present the standard cosmological models based on GR have come to be regarded as very effective models to describe the overall properties of the universe.

A. K. Raychaudhuri in his paper discusses how the standard cosmological models give a satisfactory description of the observed universe. He also points out several factors and difficulties on the basis of which some cosmologists demand a drastic revision of standard models. We thus find in his paper a balanced fare for the participants of the Workshop to form their own judgement on these issues.

The space-time geometry of the standard models is given by the Robertson–Walker metric which exhibits spherical symmetry round every point. But GR permits a vast variety of non-spherical models which have been studied and classified by Bianchi. These models are homogeneous but not isotropic. M. A. Melvin in his paper discusses certain Bianchi type universes which can describe axially-symmetric cosmologies or skew-axial cosmologies. These cosmologies have potential application in explaining large-scale electromagnetic fields pervading cosmos.

On the other hand even in standard isotropic and homogeneous cosmological models, one must be able to describe the local inhomogeneities like clusters and superclusters. Using the standard model as the background, P. C. Vaidya in his paper develops a scheme for describing the field of a rotating system. Thus, in the following articles we find a balanced presentation of some of the present-day problems of cosmology.

Some outstanding problems in cosmology

A K RAYCHAUDHURI

Physics Department, Presidency College, Calcutta 700 073, India

Controversies are familiar in science, indeed one may say that an essential ingredient in the progress of science is a conflict of ideas. Unfortunately, in cosmology, controversies linger on for years without showing any sign of resolution—this has led to a polarization where cosmologists of different schools seem to have developed a passionate attachment to particular viewpoints. Thus, on one hand there is the group, comprising perhaps the overwhelming majority of the researchers in the field, who think that standard cosmological models give an extremely satisfactory description of the observed universe—they cite the fact that while red shift observations have been extended to large values of the parameter z, the data all fit in with the Hubble diagram; besides, the microwave background radiation and the abundances of light nuclei have brought in striking support to these models from somewhat unexpected quarters. Recently, thanks to GUT, the baryon-antibaryon asymmetry has also been apparently explained in terms of the conditions of the early universe of big bang models.

There are others, however, may be a minority, who with considerable force point out numerous factors which demand some drastic modifications of standard ideas if not an outright rejection. The purpose of the present paper is to pinpoint these 'difficulties' and try to assess in as dispassionate a manner as possible how far these can be reconciled with the standard models.

It is imperative that we should spell out first of all what is meant by standard cosmology. The essential basis of standard models are the following assumptions:

(a) General relativity (GR) is the correct theory of gravitation at all stages of the universe.

While a change in the formal structure or ideas of GR at the level of Planck length and at extremely high densities may seem almost inevitable, as yet nothing definite has emerged so far and hence the assumption of validity of GR at all stages.

(b) The Riemannian geometry of the universe, as brought in by assumption (a), admits a six-parameter group of motions, the orbits of the group being three dimensional spacelike hypersurfaces. In physical language the universe is assumed to be isotropic and homogeneous—*i.e.* the cosmological principle holds good.

(c) The energy stress tensor is of the form

$$T_{ik} = (p + \rho)u_i u_k - p\, g_{ik},$$

where u^i is a unit timelike vector and $p \geq 0$, $\rho > 0$, $p \leq \rho/3$

Assumption (c) is somewhat less standard—assumptions (a) and (b) together severely restrict the form of the tensor T_{ik} so that (c) is to some extent redundant. Further one sometimes introduces a cosmological term which may be looked upon as an energy

stress of vacuum not obeying the restrictions spelt out in (c). Apparently violations of assumption (c) lies on the frontier of standard cosmology.

With these assumptions we are led to the Friedmann (or Robertson–Walker) line element:

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \frac{R^2}{(1+kr^2/4)^2}(\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2), \tag{1}$$

with $k = 0, \pm 1$, and the equations

$$\frac{8\pi\rho}{3} = \frac{k}{R^2} + \frac{\dot{R}^2}{R^2}, \tag{2}$$

$$-8\pi p = \frac{k}{R^2} + \frac{\dot{R}^2}{R^2} + \frac{2\ddot{R}}{R}. \tag{3}$$

Or,

$$\frac{\ddot{R}}{R} = -4\pi\left(p + \frac{\rho}{3}\right). \tag{4}$$

The first confrontation with observation comes from the existence of clusters and superclusters of galaxies constituting a distinct departure from the homogeneity built in the metric (1). Thus there arises a pressing need for a modified metric which on the one hand will accommodate the observed inhomogeneities and will also be consistent with (1), *i.e.* the overall homogeneity of the universe at large. Such metrics are apparently not difficult to construct—the possibility was explored by Einstein himself in a paper co-authored with Strauss. More elaborate investigations have been undertaken by a host of researchers—McVittie, Bonnor and Vaidya to name a few (Vaidya incorporates the possibility of rotation of the localised irregularities). However if one goes into details, no satisfactory solution has been found so far—thus Bonnor (1972) failed to accommodate a density variation $\rho \propto r^{-1.7}$ (given by de Vaucouleurs, 1971) for r from 10^7 to 10^{27} cm and the observed red shift-luminosity relation in any spherically symmetric metric.

Observationally, one attempts to determine $R(t)$ from the red shift formula

$$z = \frac{R_0}{R} - 1 = \left(\frac{\dot{R}}{R}\right)_0 r + \frac{1}{2}\left(\frac{\ddot{R}}{R}\right) r^2 + \ldots, \tag{5}$$

where r is the coordinate position of the emitting object and we are considered to be at the origin of coordinate system and the subscript zero indicates the epoch of observation (*i.e.* the present). For small enough values of r, it is usual to write

$$z = \left(\frac{\dot{R}}{R}\right)_0 r = H_0 r. \tag{6}$$

This linear relation, called the Hubble law, is one of the cornerstones of standard cosmology and is usually believed to be supported by observations. However recently Nicoll and Segal (1978a, b) have claimed to have obtained a better agreement with a quadratic relation $z \propto r^2$. From an analysis of the data up to $z \sim 4$, Nicoll and Segal find a satisfactory agreement with their theoretical formula (a formula derived from Segal's chronometric cosmology—a non-standard theory):

$$m = 2{\cdot}5 \log z - 2{\cdot}5(2-\alpha)\log(1+z) + M, \tag{7}$$

where m is the apparent magnitude of an object of absolute magnitude M and showing a red shift z. (α is the spectral index). For standard cosmology however the corresponding formula is

$$m = 5 \log z + 1{\cdot}086(1 - q_0)z - 5 \log H_0 - 5 + M, \tag{8}$$

where $q_0 \equiv -[(\ddot{R}/R)/(\dot{R}/R)^2]_0$ is the so-called deceleration parameter. The coefficients 2·5 and 5 of $\log z$ in (7) and (8) correspond to quadratic and linear law respectively.

If we accept, at least provisionally, the Nicoll-Segal analysis as correct, can we accommodate a quadratic law in standard cosmology? A look at (5) shows that if $H_0 = (\dot{R}/R)_0$ be vanishingly small so that the first term on the right no longer dominates, we shall have a quadratic relation. A little calculation using $H_0 = 0$ gives

$$m = 2{\cdot}5 \log z - 5 \log \left|\frac{\ddot{R}}{R}\right|_0 + M - \text{const.}$$

Thus the coefficient 2·5 is recovered and one may evaluate $(\ddot{R}/R)_0$ from the observational data. Of course whether this approach will work will depend on the value of the age of the universe that $(\ddot{R}/R)_0$ leads to. Indeed we consider the agreement, so far as the order of magnitude is concerned, between the age of the universe ($\sim H_0^{-1}$) and the estimated ages of galaxies and other astronomical bodies a remarkable result of standard cosmology. Whether this agreement will survive in the new viewpoint remains to be seen. However the following may be noted:

(a) The idea that we are at the epoch $\dot{R} \approx 0$ combined with the observed red shift directly leads to the result $k = +1$ *i.e.* the universe is closed but it may not necessitate a search for 'missing masses' if $(\ddot{R}/R)_0$ turns out to have such a value that the density determined from (4) is of the same order of magnitude as the observed density of luminous matter.

(b) There may not arise any difficulty so far as the nucleo-genesis in the early universe is concerned.

(c) If really we are at present at the epoch of maximum R, then it is a strange coincidence requiring some sort of explanation.

Leaving the Nicoll–Segal analysis, we go to another aspect of observations on red shift. Standard cosmology requires R to be a smooth function of time and hence one would expect z to vary smoothly with distance of the emitter, being uniquely determined by the value of r. According to some astronomers this is not so always—they claim to have found objects which are close together but show significantly different red shifts. Our first reaction may be that these are due to peculiar velocities but the facts of the situations make this explanation hardly tenable. One may also imagine some serious departure from homogeneity bringing in sharp peaks in the $R - t$ curve but that again is not easily reconcilable with the general theory of relativity. The protagonists of standard cosmology just dismiss this so-called anomalous red shifts as cases of wrong interpretation but that again has been firmly contradicted by the other school. To us it seems that there is a problem and no solution even going beyond standard cosmology is in sight.

Let us next consider the problems associated with the microwave background radiation. A close examination of the data has revealed that there is a large angle anisotropy which is principally of the $\cos\theta$ form (*i.e.* a first order spherical harmonic) and thus can be explained as a peculiar velocity of ourselves. One would then expect a

corresponding anisotropy effect in the red shift as well. While an anisotropy in the red shifts has indeed been observed (Rubin–Ford effect) the two anisotropies are apparently uncorrelated as the directions of the peculiar velocity calculated from the two sets of data are nearly at right angles. One may put forward a sort of explanation by hypothising abnormal peculiar velocities of our neighbouring superclusters but that seems hardly convincing. Apparently one has to construct a model which allows the red shift and the microwave background radiation to have different types of anisotropy but such models remain to be put forward.

Recent observations have revealed that there is a quadrupole component in the anisotropy of the microwave background radiation, although perhaps of smaller magnitude (Fabbri *et al* 1980; Boughn *et al* 1981)—the likely amplitude is perhaps about 0.7 mK. This would apparently require a tensor field and indeed a non-vanishing vorticity of the universe may give rise to a quadrupole term in the anisotropy. The expected amplitude in this case would be (Horak 1979)

$$T_2 = \frac{T_0 \omega_0^2 (1+z)^4}{8H_0^2(1+2q_0 z)},$$

where ω_0 is the present value of the vorticity and z corresponds to the last scattering surface of the radiation. As the value of q_0 is very much uncertain and the estimate of z also is just plausible, the above formula may not lead to any definite value of the present vorticity. But if this hypothesis is correct, the very existence of vorticity would contradict the isotropy assumption of standard cosmology and thus make the Friedmann metric inapplicable. It would also contradict the Mach principle as is commonly understood and thus rock the theories which profess to be based on that principle. Again there have been attempts to attribute the quadrupole anisotropy to local causes *e.g.* the nonhomogeneity in the mass distributions of the galctic clusters (Peebles 1981).

Some recent observations have revealed a very disturbing feature in the distribution of energy amongst different frequencies in the microwave background radiation (Woody and Richards 1979; Gush 1981). Standard cosmology considers this radiation to be a relic of the big bang which has cooled down to its present temperature due to the expansion of the universe. Thus the originally thermalised radiation should preserve its character and hence the energy distribution curve should be Planckian with possible slight departures due to the action of intervening matter like inverse Compton scattering. These would make the radiation relatively richer in high frequency quanta. However recent observations indicate that while there is the typical hump of thermalised radiation, the distribution does not fit in detail with the Planckian curve at any single temperature and more surprisingly the low frequency quanta are relatively richer. It has been suggested that the radiation is indeed a superposition of thermal radiations at two different temperatures—one of the components being the relict radiation (Alexanian and Grinstein 1980) but this explanation is hardly consistent with the ideas of standard cosmology. It is too early to say how these anomalous distribution will ultimately be understood.

The last problem from the observational side that we shall refer to comes again from the observation of red shifts. One can hope to determine the deceleration parameter q_0 from observations of red shifts if these extend to sufficiently large values of z. However there are troublesome points to be considered and for large z, the most uncertain is the

question of possible galactic evolution. One cannot thus get the value of q_0 with reliability—nevertheless the present general belief is towards a low value of q_0 and quite a number of researchers consider that the data lean towards a negative value of q_0. However from (4) with the pressure and density non-negative, a negative value of q_0 would be inconsistent with standard cosmology. The minimum violence to standard cosmology which may bring reconciliation with a negative q_0 is the re-introduction of the cosmological term. More revolutionary will be the consideration of negative energy density and/or negative pressure.

So far we have discussed the confrontation between observations and the expectations from standard cosmological models. The attitude of some of the proponents of standard cosmology is to follow Eddington's advice 'if there is a disagreement between your theory and experiment, disbelieve the experiment'. Indeed the nature of observations in cosmology gives ample scope for such an attitude. One may recall that for over two decades, cosmologists were puzzled by the discrepancy between the so-called age of the universe and the ages of the earth and other astronomical bodies and then it turned out that the astronomers have all the time proceeded on a wrong basis and thus underestimated the age of the universe by a factor of ten; all the headache was without any reason.

However there are difficulties at a theoretical level as well and these cannot be just wished away. Foremost amongst these is the singularity problem. Thanks to the very powerful theorems of Hawking and Penrose, it is now clear that this is not a problem for standard cosmology in particular but a problem of classical theory of gravitation itself. Whether a quantum theory of gravitation will give us singularity free cosmological models remains to be seen.

Another vexatious problem is a rather philosophical one and is connected with the existence of horizons in standard cosmological models. Choosing $t = 0$ at the instant of the big bang singularity, we have for any signal starting from $r = 0$ at $t = 0$ and reaching r at time t

$$\int_0^r \frac{dr}{(1 + kr^2/4)} \leqslant \int_0^t \frac{dt}{R}.$$

As $t \to 0$, let us suppose $R \sim t^\alpha$, then from (4) we get

$$4\pi\left(p + \frac{\rho}{3}\right) = -\alpha(\alpha - 1)t^{-2},$$

so that with $p, \rho \geq 0$, $0 < \alpha < 1$. Thus $\int_0^t dt/R$ converges to a finite limit and hence there can be a causal connection between two points at any time t only if these points lie at a limited distance apart. But there is an isotropy of the microwave background extending over large angles which is understandable only if those regions of the last scattering surface from which this radiation is coming had an opportunity of communication prior to the scattering of the radiation. But a simple calculation shows that these regions were beyond their horizons at that epoch. Again one way of escaping from this difficulty would be to introduce negative energy such that horizons are abolished.

Right now considerable theoretical activity is going on in exploring the possible impact of the grand unified theory (GUT) on cosmology. Indeed apparently some of the difficulties may find a solution in the light of GUT. Thus according to this theory at very high temperatures there is a spontaneous symmetry breaking which brings in an energy

stress tensor associated with vacuum. As the vacuum recognizes no preferred coordinate frame, the energy stress tensor must have a form which remains invariant under arbitrary coordinate transformations. Such a tensor has to be a scalar multiple of the Kroenecker tensor *i.e.* the vacuum energy stress tensor is of the form $T^i_k = \Lambda'\delta^i_k$—thus the vacuum supplies a term which may camouflage as a cosmological term. However unlike the cosmological term, vacuum interchanges energy with the cosmic material, hence Λ' may not be a constant but satisfy the conservation relation

$$[\Lambda'\delta^i_k + (T^i_k)_{\mathrm{mat}}]_{;i} = 0 \text{ or } \Lambda'_{,k} = -(T^i_k)_{;i}.$$

Ideally the phase change may occur sharply at the critical temperature and one would then have $\Lambda' = 0$ for $T > T_0$ and $\Lambda' \neq 0$ for $T < T_0$. However as in other phase changes of first order, there may be supercooling and a stage of metastable equilibrium and this may have very intriguing consequences in cosmology. However all these are in a highly speculative stage right now. In fact the value of Λ' is also to some extent uncertain but it appears to have a value much above what is considered to be the allowable limit from observational data. Thus Λ' may be about 10^{50} times the presently believed matter density. Thus we run into a contradiction. Just as more than half a century ago Einstein introduced a cosmological term to save cosmology from an apparent difficulty, the protagonists of GUT are proposing the artificial suggestion that besides the vacuum energy term there is also a truly cosmological term $\Lambda\delta^i_k$ with constant Λ, such that at the present epoch $\Lambda + \Lambda'$ is a small quantity whose value lies within the observationally determined bounds. However a cancellation of two large terms none of which are directly observable needs an explanation but none seem to be forthcoming so far (Recall for a moment the importance that Dirac gave to the agreement between large non-dimensional numbers) Be that as it may, if you are willing to accept the idea, then before the symmetry breaking *i.e.* in the very early universe there was no vacuum energy and hence the cosmological term $\Lambda\delta^i_k$ remained uncancelled. Thus the following evolution of scenario seems not impossible:

(i) In the earliest universe following the big bang when the radiation and ultrarelativistic particles dominate over the cosmological term, one has the usual relation $R \sim t^{1/2}$ as in standard cosmology.

(ii) In the next epoch the cosmological term dominates over the material energy tensor and one has the de Sitter behaviour $R \sim \exp \Lambda^{1/2}t$. If this stage persists for sufficiently long time, the horizons are abolished. This occurs in the stage $T > T_0$.

(iii) Below the temperature T_0, the cancellation of the true cosmological term and the vacuum energy term occurs and we have again the same behaviour as in standard cosmology.

What we would like to emphasise is that in these discussions, isotropy has invariably been assumed and the only success claimed is that a logical consistency made plausible.

References

Alexanian M and Grinstein B 1980 *Phys. Rev. Lett.* **44** 359
Barrow J D and Turner M S 1981 *Nature (London)* **292** 35
Bonnor W B 1972 *Mon. Not. R. Astr. Soc.* **159** 261
Boughn S P, Cheng E S and Wilkinson D T 1981 *Astrophys. J.* **243** L113
de Vaucauleurs G 1971 *Pub. Astrophys. Soc. Pacif.* **83** 113
Fabbri R 1980 *Phys. Rev. Lett.* **44** 1563

Guth A H 1981 *Phys. Rev. Lett.* **47** 745
Guth A H 1981 *Phys. Rev.* **D23** 347
Horak Z 1979 *Gen. Relat. Gravit.* **11** 261
Kazanas D 1980 *Astrophys. J.* **241** L59
Kolb E W and Wolfram S 1980 *Astrophys. J.* **239** 428
Mahapatra R N 1980 *Phys. Rev.* **D22** 2380
Nicoll J F and Segal I E 1978a *Proc. Nat. Acad. Sci.* (USA) **75** 535
Nicoll J F and Segal I E 1978b *Ann. Phys.* (N.Y.) **113** 1
Peebles P J E 1981 *Astrophys. J.* **243** L119
Sher M A 1980 *Phys. Rev.* **D22** 2989
Woody B D and Richards P L 1979 *Phys. Rev. Lett.* **42** 925

Axial and skew-axial homogeneous cosmologies

M A MELVIN
1300 Orchid Drive, Santa Barbara, CA 93111, U.S.A.

The assumption of "historical homogeneity"—or the persistence of spatial homogeneity in time—leads to the synchronous (time-orthogonal) form $ds^2 = -dt^2 + \gamma_{AB}(t)\,\varepsilon^A \otimes \varepsilon^B$, of the metrics for general spatially homogeneous (Bianchi) cosmologies. The ε^A are group-representation 1-forms associated with the spatial symmetry group of the given cosmology, and are independent of any particular space-time theory; in particular they do not depend on Einstein's equations and the local physics. Expressions for the ε^A in the canonical basis used by Estabrook, Wahlquist, and Behr (FWB) have been tabulated in two earlier papers (Melvin and Michalik 1980; Michalik and Melvin 1980). Explicit expressions for the scale factors $\gamma_{AB}(t)$ were also derived. These correspond to the choice of the canonical basis to describe the symmetry but, in contrast to the ε^A, they also depend on the space-time physical theory. Here the $\gamma_{AB}(t)$ come out as observable geometric-kinematic quantities (generalized Hubble constants) which also appear in the solutions of Einstein's equations for a given cosmology. Central to the discussion is the matrix $C(t)$ which relates the time-independent canonical basis describing the symmetry to a time-dependent orthonormalizing basis. With the imposed requirement that the Riemannian geometry of the evolving spatial hypersurfaces retain its quasi-canonical form in time ("quasi-canonical gauge") it has been shown that $C(t)$ is a product of a diagonal matrix D and a rotation matrix R; and a table of these forms of $C(t)$ for the various Bianchi types was given. The simple Hubble-parameter expressions for the metric scale factors $\gamma_{AB}(t)$ for all types come solely from the diagonal factor D in C. The metric, then, in all cases takes the form

$$ds^2 = -dt^2 + \gamma_{AB}\varepsilon^A \otimes \varepsilon^B = -dt^2 + (\varepsilon^{\mathrm{I}}/\alpha)^2 + (\varepsilon^{\mathrm{II}}/\beta)^2 + (\varepsilon^{\mathrm{III}}/\gamma)^2.$$

The sole three Hubble-parameters appearing are given by

$$\alpha = \exp\left(-\int s_1\,dt\right) \equiv \exp(-\sigma_1),$$
$$\beta = \exp\left(-\int s_2\,dt\right) \equiv \exp(-\sigma_2),$$
$$\gamma = \exp\left(-\int s_3\,dt\right) \equiv \exp(-\sigma_3),$$

where s_1, s_2 and s_3 are the diagonal elements of the rate-of-strain (extrinsic curvature) matrix.

There are three Lie algebra parameters a_0, b_0, c_0 or a_0, b_0, n_0 determining the symmetry type of the Bianchi cosmology. The intrinsic geometry of the spatial hypersurfaces in Bianchi cosmologies depends solely on three parameters a, b, c (non-vector type) or a, b, n (vector type) which are time evolvents of a_0, b_0, c_0 or n_0.

The equations governing the evolution of a, b, c or a, b, n are

$$\dot{a} = (s_1 - s_2 - s)a \qquad \dot{b} = (s_2 - s - s_1)b \qquad \dot{c} = (s - s_1 - s_2)c$$
$$\dot{n} = -sn \qquad (s \equiv s_3).$$

or

$$a = a_0 \exp(\sigma_1 - \sigma_2 - \sigma), b = b_0 \exp(\sigma_2 - \sigma - \sigma_1) \quad c = c_0 \exp(\sigma - \sigma_1 - \sigma_2)$$
$$n = n_0 \exp(-\sigma) \qquad (\sigma \equiv \sigma_3).$$

The Ricci curvature tensor, which represents the principal Gaussian curvatures of the spatial hypersurfaces, is given by

$$^{(3)}R_{ik} = \begin{pmatrix} R - 4n^2 + 2aA & n(b-a) & 0 \\ n(b-a) & R - 4n^2 + 2bB & 0 \\ 0 & 0 & R - 4n^2 + 2cD \end{pmatrix},$$

where the Ricci curvature scalar is

$$^{(3)}R = 6n^2 + \tfrac{1}{2}(a^2 + b^2 + c^2) - (bc + ca + ab),$$

and the abbreviations A, B and D have been introduced:

$$A \equiv \tfrac{1}{2}(b + c - a), \qquad B \equiv \tfrac{1}{2}(c + a - b), \qquad D \equiv \tfrac{1}{2}(a + b - c).$$

We follow the convention on the Riemann and Ricci *tensors which makes* them represent the *negatives* of the physical curvatures as can be seen from the geometrical interpretation of $^{(3)}R_{ik}$ originally given by Herglotz in 1916.

The terms "axial", or "skew-axial", refer to rotational symmetry, or a systematic deviation from rotational symmetry, about the third direction. Formally the axial case is obtained by setting

$$b = a \qquad (\text{implying } s_2 = s_1) \qquad \text{Axial},$$

and the Skew-axial case by setting

$$b = -a \qquad (\text{implying } s_2 = s_1) \qquad \text{Skew-axial}.$$

The possible Axial cosmologies occur in the two type-categories, non-vector and vector.

Axial Non-vector: I VII: $^{(3)}R_{ik} = 0$ Flat

$$\text{VIII, IX:} \quad -^{(3)}R_{ik} = \frac{c}{2}\begin{pmatrix} 2a - c & 0 & 0 \\ 0 & 2a - c & 0 \\ 0 & 0 & c \end{pmatrix}.$$

In type VIII, c, following c_0, is negative at all times and the two transverse physical curvatures (directions 1 and 2) are negative whereas the third is positive. In type IX c is positive and the physical Ricci curvature tensor takes the form

$$-^{(3)}R_{ik} = \tfrac{1}{2}e^{2\sigma}\begin{pmatrix} 2e^{2(\sigma_1 - \sigma)} - 1 & 0 & 0 \\ 0 & 2e^{2(\sigma_1 - \sigma)} - 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

so that as long as $c < 2a$ or $\sigma - \sigma_1 < \ln\sqrt{2}$ the physical curvatures are all positive. If the excess over longitudinal of transverse relative strain rate is greater than $\ln\sqrt{2}$ the situation is similar to that of type VIII: the transverse physical curvatures are negative while the longitudinal curvature remains positive.

The second axial category is vector axial. This can occur only in types V and VII_h. In this case the physical Ricci tensor takes the form

$$-{}^{(3)}R_{ik} = \begin{pmatrix} -2n^2 & 0 & 0 \\ 0 & -2n^2 & 0 \\ 0 & 0 & -2n^2 \end{pmatrix}$$

corresponding to uniform negative physical curvature.

The skew-axial case $b = -a$ occurs only in type VI_h (vector for $h \neq 0$, non-vector for $h = 0$) with $c = 0$ in all cases. The physical Ricci curvatures are given by

$$-{}^{(3)}R_{ik} = -\begin{pmatrix} 2n^2 & -2na & 0 \\ -2na & 2n^2 & 0 \\ 0 & 0 & 2(n^2+a^2) \end{pmatrix}.$$

This can be "super diagonalized" by a $\pi/4$ radian rotation, and it can be written:

$$-{}^{(3)}R_{ik} = -2n^2\begin{pmatrix} 1-1/n_0 & 0 & 0 \\ 0 & 1+1/n_0 & 0 \\ 0 & 0 & 1+1/n_0^2 \end{pmatrix},$$

where n_0 is a constant which can take on values from zero to infinity. In general, curvatures are negative except one transverse curvature making a 45° angle with the a and b directions. It will be positive Gaussian curvature if, $0 < n_0 < 1$.

This bisecting direction between a and b seems to have special significance also for the vector axial cases, because by rotating through $\pi/4$ about the 3-direction we can "superdiagonalize" the metrics, *i.e.* have them not only diagonal in the ε forms but also in coordinate differentials dx_1, dx_2, dx_3. The metric for the skew-axial type VI is already in this form because $\beta = \alpha$ and the spatial metric becomes

$$\left(\frac{1+k}{2n}\right)^2 (dx_1)^2 + \frac{2}{\alpha^2}\left[\exp(-2x_1)(dx_2)^2 + \exp(-2kx_1)(dx_3)^2\right]$$

$$= \left(\frac{1+k}{2}\right)^2 \exp(2\sigma)(dx_1)^2 + 2\exp(2\sigma_1)[\exp(-2x_1)(dx_2)^2$$

$$+ \exp(-2kx_1)(dx_3)^2],$$

where k is a constant parameter which is related to n_0 by

$$n_0 = (1+k)/(1-k).$$

If $k = 0$, $n_0 = 1$. Type VI becomes Type III.

The metric for Type V is also automatically in superdiagonal form. As we have stated the metric for the remaining vector axial types VII can also be brought to this form. Upon making a $\pi/4$ rotation about the x_1 axis (ϵ^{III} direction) followed by a change of scale

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}_{\mathrm{new}} = \begin{pmatrix} A & 0 & 0 \\ 0 & \sqrt{1-A^2} & 0 \\ 0 & 0 & \sqrt{1+A} \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}_{\mathrm{old}}$$

The Type VII_h spatial metric goes over to the form

$$\exp(2\sigma)(\mathrm{d}x)^2 + \exp(2\sigma_1 - 2x)(\mathrm{d}y^2 + \mathrm{d}z^2).$$

The strain rates s_1 and s, from which σ_1 and σ_2 are obtained, depend on the physical theory. In Einstein's theory, with diagonal strain rate for spatially homogeneous cosmologies the axial and skew-axial symmetric strain rates satisfy the algebraic and differential equations:

$$\text{Energy density:} \quad T_{44} = \tfrac{1}{2}(s_1 + 2s)s_1 - {}^{(3)}R/4 = \tfrac{1}{2}s_1^2 + ss_1 - (3/2)n^2.$$

$$+\frac{c}{4}\left(\frac{c}{2} - 2a\right) \text{ in all axial cases.}$$

$$\text{Momentum density:} \quad T_{43} = n(s_1 - s).$$

$$\dot{s}_1 = 2T_{11} + {}^{(3)}R_{11} - (2s_1 + s)s_1 + [-\mathrm{Tr}\,T + \tfrac{1}{2}(s_1 + 2s)s_1 - {}^{(3)}R/4],$$
$$\dot{s} = 2T_{33} + {}^{(3)}R_{33} - (2s_1 + s)s + [-\mathrm{Tr}\,T + \tfrac{1}{2}(s_1 + 2s)s_1 - {}^{(3)}R/4],$$

where $\mathrm{Tr}\,T \equiv T_{11} + T_{22} + T_{33}$, and $2T_{11} + {}^{(3)}R_{11} = 2T_{22} + {}^{(3)}R_{22}$. Since in all cases, except the superdiagonalized VI (skew-axial) case, ${}^{(3)}R_{22} = {}^{(3)}R_{11}$ we must have

$$T_{11} = T_{22} \qquad (\text{except for } VI_h).$$

For VI_h in the superdiagonalizing basis we have

$$T_{22} - T_{11} = \tfrac{1}{2}({}^{(3)}R_{11} - {}^{(3)}R_{22}) = -2n^2/n_0.$$

In VII_h and V we have

$$\dot{s}_1 = -T_{33} + n^2/2 - (3/2)s_1^2,$$
$$\dot{s} = T_{33} - 2T_{11} + n^2/2 + \tfrac{1}{2}s_1^2 - s_1 s - s^2,$$

as well as $\dot{n} = -sn$.

The solutions and metrics found earlier (Melvin and Michalik 1980; Michalik and Melvin 1980; Melvin 1975) are found by solving these equations and substituting in the general expressions for the metrics.

References

Melvin M A 1975 *Ann. N. Y. Acad. Sci.* **262** 253
Melvin M A and Michalik T R 1980 *J. Math. Phys.* **21** 1938
Michalik T R and Melvin M A 1980 *J. Math. Phys.* **21** 1952

Rotating systems in cosmological background

P C VAIDYA
Department of Mathematics, Gujarat University, Ahmedabad 380 009, India

1. Introduction

C. V. Vishveshwara, in his paper in this volume, referred to certain outstanding problems in gravitational collapse and the first problem in his list was "pre-collapse configuration with rotation". The aim in the present paper is to express some preliminary procedures by which such a configuration can be treated mathematically. The standard model for spherical gravitational collapse is given by the Oppenheimer–Snyder solution which is mathematically the same as Friedman's solution. Therefore in order to introduce the effects of rotation the first step would be to deal with a rotating system in a cosmological background (RSCB).

As cosmological background we shall take either Einstein static universe or deSitter's static universe. One can include the expanding universe background in the above scheme by using a geometry conformal to the background geometry of Einstein universe.

We shall, in §2 express background metrics in terms of what we have termed rotating spheroidal coordinates and arrive at a general form of the Riemannian metric which we will use for studying gravitational fields of RSCB. In §3 we introduce the material content of RSCB as a perfect fluid with a unidirectional flow of null radiation and derive the field equations which the g_{ik} of our metric must satisfy.

In §4 certain very general conclusions following from the field equations are derived and in the concluding section we discuss some particular cases.

2. Background metrics

Geometry of Kerr metric is rather unfamiliar and we begin with the flat background of Kerr metric. Let us start with the flat metric

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2.$$

Apply the following transformations, first given by Kerr.

$$x = \rho\cos\beta - w\sin\beta, \qquad \rho = r\sin\alpha$$
$$y = \rho\sin\beta + w\cos\beta \qquad z = r\cos\alpha \qquad w = a\sin\alpha$$

so that $(x^2+y^2)/(r^2+a^2)+(z^2/r^2) = 1$. The surfaces $r =$ constant are spheroids. We may call (r, α, β) as spheroidal polar coordinates. The flat metric now takes the form (Vaidya 1977)

$$ds^2 = dt^2 - (dr - a\sin^2\alpha\, d\beta)^2 - (r^2 + a^2\cos^2\alpha)(d\alpha^2 + \sin^2\alpha\, d\beta^2)$$

We now move from flat background to cosmological background. We know that Robertson–Walker metric is conformal to Einstein's static universe-metric, the conformal factor being a function of cosmic time. Let us therefore begin with the metric of Einstein-Universe and apply Kerr-like transformations. The Einstein universe has the metric

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2 - \frac{(xdx + ydy + zdz)^2}{R^2 - (x^2 + y^2 + z^2)}.$$

Apply the transformations

$$x = \rho\cos\beta - w\sin\beta, \quad \rho = R\sin\frac{r}{R}\sin\alpha, \qquad w = a\cos\frac{r}{R}\sin\alpha.$$
$$y = \rho\sin\beta + w\cos\beta, \quad z = R\sin\frac{r}{R}\cos\alpha,$$

The metric transforms to

$$ds^2 = dt^2 - (dr - a\sin^2\alpha\, d\beta)^2 - M^2\left[\left(1 - \frac{a^2}{R^2}\sin^2\alpha\right)^{-1} d\alpha^2 + \sin^2\alpha\, d\beta^2\right]$$

$$M^2 = (R^2 - a^2)\sin^2\frac{r}{R} + a^2\cos^2\alpha.$$

It may be noted that we have transformed coordinates only in the 3-space t = constant and have left the cosmic time t of Einstein's universe unaltered. It would be useful to introduce a null coordinate u, but we would not like to disturb the use of cosmic time as a coordinate. A null coordinate u is therefore introduced in place of r by the substitution $r = t - u$, in both cases (flat background as well as Einstein Universe). We then find that the flat background metric changes to

$$ds^2 = 2(du + a\sin^2\alpha\, d\beta)dt - (du + a\sin^2\alpha\, d\beta)^2 - (r^2 + a^2\cos^2\alpha)$$
$$\times(d\alpha^2 + \sin^2\alpha\, d\beta^2). \quad r = t - u.$$

The Einstein universe metric takes the form

$$ds^2 = 2\,(du + a\sin^2\alpha\, d\beta)dt - (du + a\sin^2\alpha\, d\beta)^2$$
$$- M^2\left[\left(1 - \frac{a^2}{R^2}\sin^2\alpha\right)^{-1} d\alpha^2 + \sin^2\alpha\, d\beta^2\right],$$

$$M^2 = (R^2 - a^2)\sin^2\frac{r}{R} + a^2\cos^2\alpha, \quad r = t - u.$$

From the above metric we can get the expanding universe background as $ds^2 = \exp[2F(t)]\, ds_E^2$, where ds_E^2 is the above Einstein universe metric, because t is the cosmic time of Einstein's universe. The coordinate labels used here are u, α, β, t. We recognise t as the cosmic time. Let us now try to get some physical meaning for the label u. For that purpose we do a little exercise with Schwarzschild's exterior solution

$$ds^2 = e^\nu dt^2 - e^{-\nu} dr^2 - r^2\,(d\alpha^2 + \sin^2\alpha)d\beta^2$$

$$e^\nu = 1 - \frac{2m}{r}, \quad m = \text{const}.$$

Introduce the retarded time u in place of t. The equation defining u is

$$u' \exp(\nu/2) + \dot{u} \exp(-\nu/2) = 0, \quad z' \equiv \frac{\partial z}{\partial r}, \quad \dot{z} \equiv \frac{\partial z}{\partial t},$$

a solution of which will be $u = t + f(r)$ with $f' = -e^{-\nu}$. Use u as a time coordinate in place of t and one gets

$$ds^2 = 2\,du\,dr + e^{\nu}\,du^2 - r^2(d\alpha^2 + \sin^2\alpha\,d\beta^2),$$

which is Eddington's form of Schwarzschild metric.

But in our scheme we use u not in place of t, but in place of r. And one can do this because u is a null coordinate. We return to a time-like coordinate T by the substitution $u = T - r$ and use (T, u, α, β) as coordinates, *i.e.*, replace r by $T - u$. Schwarzschild's metric then takes the form

$$ds^2 = 2\,du\,dT - \left(1 + \frac{2m}{r}\right) du^2 - r^2(d\alpha^2 + \sin^2\alpha\,d\beta^2), \quad r = T - u.$$

Since u was the retarded time in the original Schwarzschild metric it continues to be the same, but now it is used in place of r and so we may redesignate it as 'retarded distance'. This is the meaning of the label u which we use in our rotating systems. As a comparison with the above form of Schwarzschid metric we note that Kerr metric in the oordinates used here will take the form

$$ds^2 = 2(du + a\sin^2\alpha\,d\beta)dT - \left[1 + \frac{2mr}{r^2 + a^2\cos^2\alpha}\right](du + a\sin^2\alpha\,d\beta)^2$$
$$- (r^2 + a^2\cos^2\alpha)[d\alpha^2 + \sin^2\alpha\,d\beta^2], \quad r = T - u.$$

3. The field equations

The form of the back-ground metrics discussed above suggest that we seek to describe the gravitational field of RSCB by a general metric of the form

$$d\sigma^2 = \exp(2\varphi)\,ds^2, \tag{1}$$

$$ds^2 = 2(du + g\sin\alpha\,d\beta)dt - 2L(du + g\sin\alpha\,d\beta)^2 - M^2(d\alpha^2 + \sin^2\alpha\,d\beta^2),$$
$$\varphi = \varphi(u, t, \alpha), \; L = L(u, t, \alpha), \; M = M(u, t, \alpha), \; g = g(\alpha). \tag{2}$$

Now for the expanding universe background we know that $\varphi = \varphi(t)$ and so we can incorporate the conformal factor $\exp(2\varphi)$ in the t-coordinate and thus for considering RSCB we may put $\varphi = 0$ without any loss of generality. When we do this our metric under consideration becomes (2) which is the metric discussed in Vaidya *et al* (1976). We shall freely use some geometrical results from that paper for our purpose here. For example for the above metric (2) one uses tetrads θ^a as follows:

$$\theta^1 = du + g\sin\alpha\,d\beta, \; \theta^2 = M\,d\alpha, \; \theta^3 = M\sin\alpha\,d\beta, \; \theta^4 = dt - L\theta^1 \tag{3}$$

so that

$$ds^2 = 2\theta^1\theta^4 - (\theta^2)^2 - (\theta^3)^2.$$

Using Cartan's equations of structure, one can work out the tetrad components $R(ab)$ of the Ricci tensor. These are recorded in Vaidya *et al* (1976) and we shall use them whenever needed.

We now take the material content of RSCB as a perfect fluid with a unidirectional flow

of null radiation, the energy-momentum tensor being

$$T_{ik} = (p+\rho)v_i v_k - p g_{ik} + \sigma w_i w_k,$$
$$g_{ik} v^i v^k = 1, \quad g_{ik} w^i w^k = 0. \tag{4}$$

Einstein's equations will then connect the geometry with the material content as

$$R_{ik} - \tfrac{1}{2} g_{ik} R = -8\pi T_{ik},$$

from which one can derive the relation

$$R_{ik} = -8\pi[(p+\rho)v_i v_k - \tfrac{1}{2}(\rho - p) g_{ik} + \sigma w_i w_k] + \Lambda g_{ik}. \tag{5}$$

(We have added the cosmological constant Λ for completeness).

With the choice of tetrads (3) for the metric (2) we take the tetrad components of the vectors v_i and w_i as follows:

$$v_{(a)} = \left(\frac{\lambda}{\sqrt{2}}, 0, 0, \frac{1}{\sqrt{2\lambda}}\right), \quad w_a = (1, 0, 0, 0)$$

which will satisfy the conditions (4). We then find from (5) that

$$\left.\begin{aligned} R_{(11)} &= -8\pi\left[\frac{\lambda^2}{2}(p+\rho)+\sigma\right] \\ R_{(22)} &= -8\pi\left[\frac{1}{2}(\rho - p)\right] = R_{(33)} \\ R_{(44)} &= -8\pi\left[\frac{1}{2\lambda^2}(p+\rho)\right] \\ R_{(14)} &= -8\pi p \end{aligned}\right\} \tag{6}$$

$$R_{(23)} = R_{(24)} = R_{(34)} = R_{(12)} = R_{(13)} = 0. \tag{7}$$

The geometrical forms of the components $R_{(ab)}$ are recorded in Vaidya *et al* (1976). It is seen there that $R_{(22)} = R_{(33)}$ and $R_{(23)} = 0$ are satisfied identically. Then the four equations (6) would give us the four physical variables p, ρ, σ and λ and the four equations (7) *viz.*

$$R_{(24)} = 0, \quad R_{(34)} = 0, \quad R_{(12)} = 0, \quad R_{(13)} = 0,$$

give us the differential equations to determine the geometry through the functions L, M and g giving g_{ik}. Borrowing the components of $R_{(ab)}$ from Vaidya *et al* (1976) our field equations written explicitly are

$$\begin{aligned} &(M_t/M)_y - (f/M^2)_u = 0, \qquad g\,\mathrm{d}\alpha = \mathrm{d}y, \\ &(M_t/M)_u + (f/M^2)_y = 0, \qquad 2f = g_\alpha + g\cot\alpha, \end{aligned} \tag{8}$$

$$\begin{aligned} &(L_t + M_u/M)_y + (2Lf/M^2)_u = 0, \\ &(L_t + M_u/M)_u - (2Lf/M^2)_y = 0; \end{aligned} \tag{9}$$

(a suffix denotes partial derivative with respect to the corresponding variable: thus $Z_x \equiv \partial Z/\partial x$ etc.).

4. A general solution

The equations $R_{(42)} = 0$, $R_{(43)} = 0$ are the two equations (8) above and it will be seen that they are partial differential equations for a single function $Z = f/M^2$. The two equations are equivalent to

$$\theta_y = U\sin\theta + (U_u/U);\ \theta_u = -U\cos\theta - (U_y/U), \tag{10}$$

where $\theta_t = 2Z$ and $U_t = 0$, $U = U(u, y)$. The condition of consistency of (10) gives an equation for U

$$\nabla^2 \log U = U^2, \quad \nabla^2 = \frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial y^2}. \tag{11}$$

Thus to find $Z = f/M^2$ satisfying $R_{(42)} = 0$, $R_{(43)} = 0$, one must first solve the U-equation (11) and then solve any one of the two θ-equations to get θ containing an undetermined function of t. Then $\theta_t = 2Z$ determines Z.

One can similarly solve the two equations (9) giving $R_{(12)} = 0$, $R_{(13)} = 0$ and write down an explicit form for the function L satisfying them. However it is possible to prove some general theorems without explicitly solving (9). One can prove the following theorems easily.

Theorem 1. If $R_{(42)} = 0$, $R_{(43)} = 0$ then $R_{(44)}$ will be a constant or at most a function of t only.

Theorem 2. If $R_{(42)} = 0$, $R_{(43)} = 0$ and in addition $R_{(12)} = 0$, $R_{(13)} = 0$ then $R_{(14)} - LR_{(44)}$ will be a constant or at most a function of t only.

These theorems increase our faith that the rotating systems described by this scheme will have a cosmological background.

An explicit form for the function L satisfying (9) can be expressed as

$$2L = at\, U\sin\theta + b\, U\cos\theta - \dot{m} + 2m(\dot{M}/M), \tag{12}$$

where an overhead dot indicates differentiation with regard to t, m is an undetermined function of t and a and b are undetermined constants. The above is not most general form of L satisfying (9), but it is general in the sense that we are not using any particular form of $Z = f/M^2$. Thus (12) satisfies (9) whatever U satisfying (11) and θ satisfying (10). For an explicit solution (U, θ) of (11) and (10) one can work out M^2 and then get $2L$ satisfying (9) which may not be a particular case of (12). Since our aim is to present a workable scheme we do not go into the details.

In the next section we take one such simple particular case to illustrate how the physical aspects of the system can be worked out.

5. A simple particular case

$U = 1/y$ is a simple solution of the U-equation (11). Then the first of the two θ-equations (10) will lead to a simple solution $\tan\theta/2 = y/r$ where r is an undetermined function of u and t. We choose $r = t - u$. Then $\frac{1}{2}\theta_t = -y/(r^2 + y^2)$ giving

$$Z = \tfrac{1}{2}\theta_t = -y/(r^2 + y^2) \quad \text{or} \quad M^2 = (f/-y)(r^2 + y^2) \tag{13}$$

One can now take $L = L(r, y)$. Then it is a straight-forward matter to solve $R_{(12)} = 0$

$R_{(13)} = 0$, to get

$$2L = 1 + [2mr/(r^2+y^2)] + k(r^2+y^2), \tag{14}$$

m and k being undetermined constants. (13) and (14) give the two functions M and L satisfying the field equations. The function $g(\alpha)$ appearing through $f = f(y)$ has remained undetermined upto this stage. Using these explicit forms of M and L one can write down explicitly the geometric components $R_{(ab)}$, one will then find

$$R_{(44)} = 0,\ R_{(14)} = 3k,\ R_{(22)} = -3k + [(2+4ky^2 - yG)|(r^2+y^2)]$$
$$R_{(11)} = -2yk[(g^2/f)+2y]|(r^2+y^2).$$

In the above we have written G as defined by

$$2fG = g^2[(1/y^2)+(f_y/f)_y] + 2f_y - 2.$$

Using equations (6) connecting $R_{(ab)}$ with physical parameters we find that

$$8\pi p = -3k+\Lambda, 8\pi\rho = 3k-\Lambda, 8\pi\sigma = 2yk[(g^2/f)+2y]|(r^2+y^2)$$

along with $Gy - 2 - 4ky^2 = 0$.

It is clear that we must choose $k = \frac{1}{3}\Lambda$ so that

$$p = 0,\ \rho = 0,\ 8\pi\sigma = \tfrac{2}{3}y\Lambda[(g^2/f)+2y]|(r^2+y^2)$$
$$Gy - 2 - \tfrac{4}{3}\Lambda y^2 = 0. \tag{15}$$

The last equation (15) will determine the function $g(\alpha)$.

If $\Lambda = 0$, (15) will be satisfied if $f = -y$ *i.e.*, if $g = a\sin\alpha$, a being a constant, and we recover the Kerr metric.

The other simple case is $\Lambda \neq 0$, $m = 0$. Then

$$2L = 1 + \frac{\Lambda}{3}(r^2+y^2) = 1 + \frac{r^2+y^2}{R^2} \quad \text{if} \quad \frac{1}{R^2} = \frac{\Lambda}{3},$$
$$M^2 = (f/-y)(r^2+y^2), p = 0, \rho = 0, 8\pi\sigma = 2yR^2[(g^2/f)+2y]|(r^2+y^2).$$

where $g = g(\alpha)$ or $f = f(y)$ satisfies

$$Gy - 2 - \frac{4y^2}{R^2} = 0. \tag{16}$$

One can interpret this simple solution as a rotating DeSittar metric because (i) it represents a universe devoid of fluid ($p = 0, \rho = 0$), (ii) it reduces to the usual DeSitter metric when the rotation parameter $g = 0$. It has an additional feature that the universe is pervaded by a unidirectional flow of null radiation which arises solely due to rotation. As a matter of fact if one solves (16) correctly upto the 1st power of $1/R^2$ one can show that

$$8\pi\sigma = \frac{2a^2}{R^2}\frac{(3\cos^2\alpha - 1)}{r^2 + a^2\cos^2\alpha},$$

so that $\displaystyle\int_0^{\pi/2}\int_0^{2\pi} M^2\sigma\sin\alpha\, d\alpha\, d\beta = 0,$

i.e. the net out-flow of null radiation across the 2-space with metric $M^2(d\alpha^2 + \sin^2\alpha\, d\beta^2)$ is zero. Thus there is no net loss of energy due to this flowing radiation. The curved

nature of space-time and the rotation introduced in it, together, so to say, lead to a churning of the gravitational energy in the DeSitter space-time which flows out from a cone of semi-angle arc cos $(1/\sqrt{3})$ and with the axis coinciding with the axis of rotation and returns through the cone of semi-angle arc sin $(1/\sqrt{3})$ with a perpendicular axis.

References

Vaidya P C 1977 *Pramana* **8** 512

Vaidya P C, Patel L K and Bhatt P V 1976 *General Rel. Grav.* **7** 701

§ V QUANTUM COSMOLOGY

INTRODUCTION

To the uninitiated, the title quantum cosmology may appear puzzling. Quantum theory is usually applied to microscopic systems whereas cosmology deals with the large-scale structure of the universe. How then can the two subjects come together in a coherent way?

The papers presented in this section answer this question and attempt to describe some of the investigations which form part of the developing subject of quantum cosmology. Classical gravity as given by Einstein's general relativity leads us to the well-known big-bang cosmologies. Attempts at grand unification take us to the early history of the universe when it was $\sim 10^{-36}$ second old (or young). As we trace the history of the universe we find that classical gravity which is reliable right down to such early epochs, breaks down before the big bang instant of $t = 0$ is reached. Quantum considerations become relevant at $t \sim 10^{-44}$ second.

To what extent are quantum considerations important to cosmology? Is the classical deduction of the existence of big bang singularity consistent with quantum ideas? Do stationary states of the universe exist? What techniques are useful in arriving at quantitative conclusions about the quantum universe? Such questions are discussed in the papers that follow.

Quantum fluctuations near the classical space-time singularity

JAYANT V NARLIKAR

Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

1. Approach to quantum gravity

Gravity is one of the four basic interactions of physics known today. The other three interactions are the strong interaction, the weak interaction and the electromagnetic interaction. Of these the first two are of short range, *i.e.* they are effective over sub-atomic distances of the order 10^{-12} cm. For their discussion one is necessarily driven to the use of quantum framework. The electromagnetic interaction is of long range and its effects are felt at macroscopic level. The successes of the electromagnetic theory at the classical level were not sufficient, however, for the full understanding of its implications. Quantum electrodynamics was necessary for understanding several phenomena at the atomic level.

Take for example, the hydrogen atom. The electron circulating round the proton in the H-atom should not stay in the same orbit if the classical Maxwell–Lorentz electrodynamics is any guide. The classical electron is expected to circulate round in steadily smaller orbits, ultimately spiralling into the proton. The characteristic time-scale for this to happen is very small.

$$\tau \sim \frac{e^2}{m_e c^3} \sim 10^{-23}\,\mathrm{s}. \tag{1}$$

The quantum theory of the H-atom solves the problem satisfactorily. Instead of the electron moving in a continuously changing orbit (as the classical theory requires) it goes in one of the many discrete orbits. These orbits are *stationary* and the electron can continue moving along a stationary orbit unless it is disturbed by outside agency or unless it makes a spontaneous transition to an orbit of lower energy. Quantum theory tells us that even the lowest energy orbit is of finite size. The characteristic linear size to emerge from the quantum calculations was

$$a \sim \frac{\hbar^2}{m_e e^2} \sim 10^{-8}\,\mathrm{cm}. \tag{2}$$

More sophisticated calculations show further small-scale effects such as the Lamb shift which have been confirmed by experiments. Thus by now quantum electrodynamics is considered a well-established, indeed the best understood, part of physics.

What about gravity? Like electromagnetic theory, gravity is a long-range interaction. Can it be quantized? Does it have to be quantized? If so what new effects may we expect from quantum gravity? Can these effects be measured by laboratory experiments as the effects of electromagnetic theory were measured and established?

Leaving the first question aside for the time being we will turn our attention to the later questions. Our guidelines from the rest of physics suggest that gravity should also be discussed within the quantum framework. A rule of the thumb for deciding when the

quantum framework is relevant is the one suggested by Dirac (1935). This is as follows.

Consider the classical action $\mathscr{A}$ describing any physical interaction. The classical behaviour of the interaction is given by the principle of stationary action

$$\delta \mathscr{A} = 0. \tag{3}$$

For such a theory the quantum considerations become important if in the relevant space-time region $\mathscr{V}$ the quantity of action becomes *comparable* to or *less* than $\hbar$:

$$\mathscr{A} \lesssim \hbar. \tag{4}$$

Let us apply these considerations to the Einstein–Hilbert action of general relativity

$$\mathscr{A} = \frac{c^4}{16\pi G} \int_{\mathscr{V}} R\sqrt{-g}\, \mathrm{d}^4 x. \tag{5}$$

Suppose in a spherical region of radial size. L the average density is ρ. Taking the time interval to be L/c we get for $\mathscr{A}$,

$$\mathscr{A} \sim \frac{c^4}{16\pi G} \cdot \frac{8\pi G}{c^4} \rho c^2 \cdot \frac{4\pi}{3} L^3 \cdot \frac{L}{c} = \frac{2\pi}{3} \rho L^4 c, \tag{6}$$

where we have used the Einstein equations to estimate R in terms of ρ. In a case of strong gravity we would have the black hole limit:

$$L \gtrsim \frac{2G}{c^2} \cdot \frac{4\pi}{3} \rho L^3,$$

i.e. $$\rho L^2 \lesssim 3c^2/8\pi G. \tag{7}$$

From (6) and (7) we get

$$\mathscr{A} \lesssim L^2 c^3/4G. \tag{8}$$

Therefore the quantum condition (4) becomes an upper limit on the linear size

$$L \sim 2(G\hbar/c^3)^{1/2} \sim 2L_p, \tag{9}$$

where $$L_p = (G\hbar/c^3)^{1/2} \tag{10}$$

is called the Planck length. Since G, $\hbar$ and c are the only fundamental constants available to a quantum theory of gravity the Planck length inevitably turns up in any approach to quantum gravity.

The smallness of L_p suggests that the quantum effects of gravity are not important at a microscopic level. This explains why unlike quantum electrodynamics, quantum gravity has not come up with any laboratory experiment. Indeed there is no gravitational phenomenon known today which would cease to exist if $\hbar$ were zero. Contrast this situation with that in electrodynamics. If $\hbar$ were zero, there would be no spectroscopy, no Compton effect, no photoelectric effect . . . .

Why then concern ourselves with a subject which has no demonstrable effect? The justification for studying quantum gravity can be given on two grounds. (i) The *aesthetic* reason that for completeness we should develop the quantized version of gravity. After all the approach to unifying different physical interactions is also guided by an aesthetic motivation, for the so-called grand unified theories are really tested at energies far beyond those of any man-made accelerators. (ii) High energy astrophysics and

cosmology may provide scenarios where quantum gravity becomes relevant. We will discuss two such scenarios in this paper.

Having established the 'why', the next question of 'how' is more difficult to answer. Several different approaches to quantum gravity exist, none of them being entirely satisfactory. In this paper we adopt the path integral approach which has proved useful in other areas of quantum theory. We begin with a brief description of this approach.

2. The path integral formalism

Let us consider a simple problem in classical mechanics, the problem of the free particle moving in one dimension. The action describing this motion in the Newtonian framework is

$$\mathscr{A} = -\int \tfrac{1}{2} m\dot{x}^2 \, dt. \tag{11}$$

Here m is the mass of the particle, x its displacement from origin at time t. $\delta\mathscr{A} = 0$ gives us the equation of motion

$$m\ddot{x} = 0. \tag{12}$$

If we are given that the particle was at $P_1(x_1, t_1)$ to start with and at $P_2(x_2, t_2)$ to end with, the path of the particle is uniquely fixed as one given by the equation

$$\bar{x}(t) = x_1 + \frac{x_2 - x_1}{t_2 - t_1}(t - t_1), \tag{13}$$

for $t_1 \leq t \leq t_2$. Denote this path in the space-time diagram by $\bar{\Gamma}$, and refer to it as the *classical path.*

A general path from P_1 to P_2 may be denoted by Γ, and it is given by a C^2 function $x(t)$ with $x(t_1) = x_1$ and $x(t_2) = x_2$. The action evaluated along Γ is therefore a functional of $x(t)$. We will denote it by $\mathscr{A}(\Gamma)$. Thus, we have

$$\bar{\mathscr{A}} \equiv \mathscr{A}(\bar{\Gamma}) = -\frac{m(x_2 - x_1)^2}{2(t_2 - t_1)}. \tag{14}$$

It was Feynman (1948) who gave a prescription which relates classical mechanics to quantum mechanics and which gives a more precise expression to the Dirac's concept described earlier. The Feynman rule is to attach a probability amplitude $\mathscr{P}(\Gamma)$ to each path Γ from P_1 to P_2 and to define a propagator $K[P_2; P_1]$ as a sum of $\mathscr{P}(\Gamma)$ over all paths Γ. Thus we have

$$\mathscr{P}(\Gamma) = \exp\left\{\frac{i\mathscr{A}(\Gamma)}{\hbar}\right\}, \tag{15}$$

$$K[P_2; P_1] = \sum_{\Gamma} \mathscr{P}(\Gamma). \tag{16}$$

The sum over paths is in practice an integral over the continuum of all paths. In the particular case of the free particle the path integral can be explicitly evaluated to give for $t_2 > t_1$

$$K[x_2, t_2; x_1, t_1] = \left(\frac{im}{2\hbar(t_2 - t_1)\pi}\right)^{1/2} \exp\left\{-\frac{im(x_2 - x_1)^2}{2\hbar(t_2 - t_1)}\right\}. \tag{17}$$

$K = 0$ for $t_2 < t_1$. Thus all paths are supposed to be future directed.

The two-point function gives the probability amplitude for the particle to be at P_2 given that it was at P_1. Quantum mechanically speaking we no longer have a unique path to go from P_1 to P_2. All paths Γ contribute towards K which describes the overall effect of interference of probability amplitudes along the different paths. The fact that we cannot definitely say which path the particle took from P_1 to P_2 is indicative of the lack of determinism in the quantum framework.

The quantum uncertainty principle tells us that we cannot even assert definitely that the particle was at P_1 or that it will be at P_2. All we can say is that there was a probability amplitude $\psi(x_1, t_1)dx_1$ that it was in the range $(x_1, x_1 + dx_1)$ at t_1 and that similarly $\psi(x_2, t_2)dx_2$ describes the probability amplitude for the particle to be in the range $(x_2, x_2 + dx_2)$ at t_2. The functions ψ are called wavefunctions and they are related by the propagator K as follows:

$$\psi(x_2, t_2) = \int_{-\infty}^{\infty} K[x_2, t_2; x_1, t_1] \psi(x_1, t_1) dx_1. \tag{18}$$

This interpretation follows directly from Feynman's rule of sum over paths.

Let us consider now the 'classical limit' implied by $\hbar \to 0$. In this limit the exponential in (15) oscillates on the unit circle $|\mathscr{P}| = 1$ with great rapidity so that with the exception of a few paths the contributions of all paths tend to cancel. The exceptional paths lie in the neighbourhood of the classical path $\bar{\Gamma}$ along which $\delta\mathscr{A} = 0$. For, by virtue of the stationarity of action the paths near $\bar{\Gamma}$ contribute the same value of the exponential and these contributions add coherently to give

$$K \to \bar{K} \sim \exp\{i\mathscr{A}(\bar{\Gamma})/\hbar\}, \tag{19}$$

as $\hbar \to 0$. This explains why the classical principle of least action holds in the classical limit.

Consider next the application of (18) to the stationary wavepacket. We may idealize the wavepacket by the wavefunction

$$\psi(x_1, t_1 | \Delta_1) = (2\pi\Delta_1^2)^{-1/4} \exp\left\{-\frac{x_1^2}{4\Delta_1^2}\right\}, \tag{20}$$

where Δ_1 is a constant showing the characteristic spread of the wavepacket at $t = t_1$. The probability of finding the particle at $x \in [x_1, x_1 + dx_1]$ is given by the Gaussian

$$|\psi(x_1, t_1 | \Delta_1)|^2 \, dx_1 = (2\pi\Delta_1^2)^{-1/2} \exp\left\{-\frac{x_1^2}{2\Delta_1^2}\right\} dx_1, \tag{21}$$

with the mean position at the origin $x_1 = 0$. We will have occasion to use such a Gaussian wavepacket in later work.

Applying (18) to (20) gives us another wavepacket at t_2. The probability in (21) is then given by

$$|\psi(x_2, t_2 | \Delta_2)|^2 \, dx_2 = (2\pi\Delta_2^2)^{-1/2} \exp\left\{-\frac{x_2^2}{2\Delta_2^2}\right\} dx_2, \tag{22}$$

where $$\Delta_2^2 = \Delta_1^2 + \frac{\hbar^2 (t_2 - t_1)^2}{4m^2 \Delta_1^2}. \tag{23}$$

Thus (23) shows that the wavepacket steadily spreads around the mean value $x_2 = 0$. In other words, the quantum uncertainty grows with time albeit at a steady rate.

We will find the above concepts useful when considering the role of quantum uncertainty in the description of space-time geometry. The above example from mechanics can evidently be generalized and in principle applied to any system describable by the action principle. Such applications may be found in the book by Feynmann and Hibbs (1965).

Two results from the path integral theory will be used in our applications to quantum gravity. The first result gives a generalization of the simple example described above and is stated as follows. Suppose the action is of the following form:

$$\mathscr{A} = \int_{t_1}^{t_2} [\alpha(t)\dot{x}^2 + 2\{\beta(t) + \gamma(t)x\}\dot{x} + \lambda(t)x^2 + 2\mu(t)x + v(t)]dt \tag{24}$$

and suppose that the solution of $\delta\mathscr{A} = 0$ gives us the classical path

$$\bar{\Gamma}: x = \bar{x}(t). \tag{25}$$

It then follows that the propagator is given by

$$K[x_2, t_2; x_1, t_1] = f(t_1, t_2)\exp\{i\mathscr{A}[\bar{x}(t)]/\hbar\}. \tag{26}$$

Thus so long as the action is quadratic in $x, \dot{x}$ the resulting path integral is explicitly solvable. The function $f(t_1, t_2)$ is usually determinable from the transitive property of the propagator:

$$K[x_3, t_3; x_1, t_1] = \int_{-\infty}^{\infty} K[x_3, t_3; x_2, t_2]K[x_2, t_2; x_1, t_1]dx_2. \tag{27}$$

The second result is the relationship between the path integration and the Schrödinger equation. If we introduce potential functions in the action and construct a hamiltonian $H(x, p, t)$ for the position (x) and momentum (p) variables then the function K satisfies the Schrödinger equation

$$\left[i\hbar\frac{\partial}{\partial t_2} - H\left(x_2, -i\hbar\frac{\partial}{\partial x_2}, t_2\right)\right]K[x_2, t_2; x_1, t_1] = \delta(x_2 - x_1). \tag{28}$$

The wavefunction ψ satisfies the source-free form of the above equation. The stationary states are given by the eigen-states of the energy operator $i\hbar\partial/\partial t$:

$$i\hbar\frac{\partial}{\partial t_2}\psi(x_2, t_2) = E\psi(x_2, t_2). \tag{29}$$

The stationary state with the wavefunction $\phi(x_2)\exp(-iEt_2/\hbar)$ has the energy E. The function ϕ satisfies the differential equation

$$H\phi = E\phi \tag{30}$$

3. The thin sandwich and superspace

The classical equations of general relativity are derived from the action

$$\mathscr{A} = \frac{c^4}{16\pi G}\int_{\mathscr{V}} R\sqrt{-g}\, d^4x + \mathscr{A}_m. \tag{31}$$

where R = scalar curvature and $\mathcal{A}_m$ is the action describing the matter present in the space-time region $\mathcal{V}$. The Einstein equations

$$R_{ik} - \frac{1}{2} g_{ik} R = -\frac{8\pi G}{c^4} T_{ik}, \tag{32}$$

follow from the variation $g_{ik} \to g_{ik} + \delta g_{ik}$ of the space-time metric tensor.

We will denote the classical solution of these equations with prescribed boundary conditions by the metric tensor $\bar{g}_{ik}$. How can we bring path integration into this picture? What are the 'points' which these paths are supposed to connect?

The points are 3-geometries ${}^{(3)}\mathcal{G}$ in an abstract 'superspace'. To understand the concept imagine a region $\mathcal{V}$ of space-time sandwiched between two space-like hypersurfaces Σ_1 and Σ_2. If we consider an arbitrary space-like hypersurface Σ between Σ_1 and Σ_2, we can describe by ${}^{(3)}\mathcal{G}$ the geometry on it, the geometry being that of a 3-space. To get the total picture we need the intrinsic and the extrinsic curvatures at any point on Σ. Alternatively, given two 3-geometries, ${}^{(3)}\mathcal{G}_1$, on Σ_1 and ${}^{(3)}\mathcal{G}_2$ on Σ_2 the Einstein equations determine (supposedly uniquely) the sequence of 3-geometries from Σ_1 to Σ_2 along such intermediate space-like hypersurfaces as Σ. For a full description of the boundary value problem see Misner *et al* (1973).

Quantization consists of imagining other sequences than the classical one, which start from a given ${}^{(3)}\mathcal{G}_1$ on Σ_1 and end with a given ${}^{(3)}\mathcal{G}_2$ on Σ_2. Each sequence may be called a 'path' Γ and the classical sequence called the path $\bar{\Gamma}$. The amplitude along each path is then given by Feynman's prescription and we end up with the propagator

$$K[{}^{(3)}\mathcal{G}_2, \Sigma_2, {}^{(3)}\mathcal{G}_1, \Sigma_1] = \sum_{\Gamma} \exp[i\mathcal{A}[\Gamma]/\hbar], \tag{33}$$

describing the probability amplitude of arriving at the 3-geometry ${}^{(3)}\mathcal{G}_2$ on Σ_2, starting from a 3-geometry ${}^{(3)}\mathcal{G}_1$ on Σ_1.

The classical problem posed above and its quantum counterpart suffer from one serious defect. The way of specifying 3-geometries as initial and final 'states' does not always lead to a solution or, when it does so, the uniqueness is not guaranteed. It was pointed out by Isenberg and Wheeler (1979) that a more satisfactory specification of the conditions on Σ_1 and Σ_2 is in the form of the conformal part of the 3-geometry and the trace of the extrinsic curvature.

To understand what this means consider any two infinitesimal line segments PA, PB drawn in Σ_1 (or Σ_2) at any point P. The specification of the conformal part of 3-geometry tells us the angle APB and the ratio of the lengths $PA:PB$. The absolute scale for length is *not* specified.

The trace of the extrinsic curvature specifies the rate of change of the scale factor at every point of the hypersurface. The classical problem is specified uniquely in terms of these initial (and final) values. In the corresponding quantum problem, the propagator K will be accordingly defined as a sum over 'paths'.

Although the above prescription is concisely started, it is not easy to translate into practice. How do we sum over paths in superspace? What is the relevant measure? Since these questions are largely unresolved, we will adopt a more modest aim in our approach to quantization.

4. Conformal fluctuations

From a given classical solution generate another geometry by a conformal transformation

$$g_{ik} = \Omega^2 \bar{g}_{ik}, \tag{34}$$

where Ω is a general function of space and time. Writing $\Omega_i = \partial\Omega/\partial x^i$ where x^i are the co-ordinates and using $\bar{g}_{ik}$ to raise or lower suffices we get

$$\int_{\mathscr{V}} R\sqrt{-g}\,d^4x = \int_{\mathscr{V}} (\Omega^2\bar{R} - 6\Omega_i\Omega^i)\{-\bar{g}\}^{1/2}\,d^4x + \int_{\partial\mathscr{V}} 6\Omega\Omega^i\{-\bar{g}\}^{1/2}\,d\Sigma_i. \tag{35}$$

Here the surface integral has appeared because R contains second derivatives of the metric tensor. Normally the action is supposed to contain only the first derivatives of the dynamical variables. To 'remove' the unwanted derivatives Hawking and Gibbons (1978) suggested that an extra surface term be added to the Hilbert action, a term whose variation would cancel out the 'unwanted' surface term of (35). Thus henceforth we ignore the surface integral over $\partial\mathscr{V}$.

Notice that the action (35) is quadratic over Ω, a circumstance which makes it relatively easy [in view of its similarity to (24)] to write down the propagator explicitly. The following example is given:

In Friedmann cosmology, near the classical big bang the dust solution is of the form

$$d\bar{s}^2 = c^2dt^2 - (t/t_0)^{4/3}[dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)], \tag{36}$$

where t_0 is a constant. The density of matter is given by

$$\rho = \rho_0\frac{t_0^2}{t^2}, \qquad \rho_0 = \text{constant}. \tag{37}$$

In this universe consider a region of co-ordinate volume V, and write the conformal *departure* from the classical solution by

$$\Omega - 1 = \Phi. \tag{38}$$

The propagator now describes the probability amplitude to go from a state of $\Phi = \Phi_1$ on $\Sigma = \Sigma_1$ to a state of $\Phi = \Phi_2$ on $\Sigma = \Sigma_2$. How are the surfaces Σ_1, Σ_2 denoted?

Of course the Weyl hypersurfaces t = constant are the natural choices for the family Σ. It is convenient, however, to define a new time co-ordinate τ by

$$\tau = t^{1/3}. \tag{39}$$

The propagator is then given by

$$K[\Phi_2, \tau_2; \Phi_1, \tau_1] = \left[\frac{3iV\rho_0\tau_1^2\tau_2^2}{4\pi h(\tau_1 - \tau_2)}\right]^{1/2} \exp iQ, \tag{40}$$

where $$Q = \frac{3V\rho_0 c^2}{4h(\tau_1 - \tau_2)}\{\tau_1^3(\tau_1 - 2\tau_2)\Phi_1^2 + 2\tau_1^2\tau_2^2\Phi_1\Phi_2 + (\tau_2 - 2\tau_1)\tau_2^3\Phi_2^2\} \tag{41}$$

For details see Narlikar (1978).

We will assume $\tau_2 < \tau_1$ and study the implications of this as $\tau_2 \to 0$, the so-called classical big bang. Using the wavepacket (20) and the relation (18) for the above propagator leads to the following analogue of (23):

$$\Delta_2 = \frac{\hbar}{3V\rho_0 c^2 \tau_1 \Delta_1} \left\{ 1 + \left(\frac{3V\rho_0 \Delta_1^2 c^2}{\hbar} \tau_1^3 \right)^2 \right\}^{1/2} \tau_2^{-2}. \tag{42}$$

Notice that Δ_2 diverges as $\tau_2 \to 0$. Our stationary wavepacket therefore becomes infinitely dispersed as the classical singularity is approached. In other words, the quantum uncertainty becomes so large that the classical solution ceases to have any significance.

This result was advanced as a conjecture by Hoyle and Narlikar (1970). More specifically these authors had argued that the nonclassical cosmologies ($\Gamma \neq \bar{\Gamma}$) may permit horizonless models. The presence of a particle horizon in the classical Friedmann cosmology has always been something of an embarrassment. For, a particle horizon inhibits the transmission of physical signals over arbitrarily large distances and therefore, it is not possible to argue that the homogeneity presently observed in the Universe arose from physical processes, just as one argues that the early Universe reached thermodynamic equilibrium through frequent scatterings. Attempts to obtain horizonless cosmologies within the classical framework have failed. The above quantum treatment shows that such models can come out of quantum cosmology with finite probability.

The above result can be generalized to include the conformal fluctuations of a general space-time manifold containing dust (see Narlikar 1981). It can be shown that the conformal fluctuations with Ω a function of all four x^i, diverge at the classical space-time singularity provided the Green's function of the scalar wave operator

$$\Box + \tfrac{1}{6}R. \tag{43}$$

diverges at the singularity. Plausible arguments and explicit demonstrations in specific (but fairly general) scenarios show that such a divergence does occur. Therefore, we have every reason to believe that quantum fluctuations diverge at the classical space-time singularity.

References

Dirac P A M 1935 *The principles of quantum mechanics* (Oxford: Clarendon Press)
Feynman R P 1948 *Rev. Mod. Phys.* **20** 367
Feynman R P and Hibbs A R 1965 *Quantum mechanics and path integrals* (New York: McGraw Hill)
Hawking S W and Gibbons G W 1978 *Phys. Rev.* **D15** 2752
Hoyle F and Narlikar J V 1970 *Nature (London)* **228** 544
Isenberg J and Wheeler J A 1979 in *Relativity, quanta and cosmology* (ed.) De Finis (New York: Johnson Reprint Corpn)
Misner C W, Thorne K S and Wheeler J A 1973 *Gravitation* (San Francisco: Freeman)
Narlikar J V 1978 *Mon. Not. R. Astr. Soc.* **183** 159
Narlikar J V 1981 *Found. Phys.* **11** 473

Discussion

B. R. Iyer: (1) If one considered the nonconformal fluctuation could there be a cancellation of the divergences? (2) Within this formalism what is the 'Vacuum' state in

particle physics sense? Can you say something on its transformation under coordinate or conformal transformations?

J. V. Narlikar: (1) I would be very surprised if such a cancellation were to occur. (2) In a certain sense the conformal function behaves like a scalar field in curved space time.

N. Dadhich: In the stellar collapse, will not event horizon and global hyperbolicity etc. make things difficult?

J. V. Narlikar: They might do. If, however, one limits the space-time to the interior of the collapsing dust ball, there are no such difficulties.

Stationary states in quantum cosmology

T PADMANABHAN

Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005. India.

1. Concept of singularity-classical vs quantum

Classical Einstein equations lead to singularities (Hawking and Ellis 1973). The 'singularity theorems', proved in the sixties, acclaiming the existence of singularities under very general considerations, have robbed physics of its predictive power.

Classically, singularity is an event with infinitely large gravitational forces. Can one trust classical gravity under these bizarre conditions? An elementary dimensional analysis shows that quantum effects are important when gravity varies over regions of the size of Planck length,

$$L_p = (G\hbar/c^3)^{1/2}. \tag{1}$$

In other words, one has to re-examine the classical singularity theorems from a quantum mechanical point of view.

How can this be done? Admittedly, we do not have today a theory for quantum gravity. Thus it is imperative that we attack the problem in two steps: First of all one should see how the quantum fluctuations affect the classical picture given above. One should then produce an alternative (quantum description) to the classical scenario.

Both these stages can be accomplished by treating the conformal part of the spacetime as a quantum variable. This method, elucidated in the paper by Narlikar (1983) (hereafter referred to as I), involves computing the path integral kernel,

$$K[\Omega_2 t_2;\Omega_1 t_1] = \int \mathscr{D}\Omega \exp \frac{i}{\hbar} \mathscr{A}[\Omega] \tag{2}$$

using the expression for the action,

$$\mathscr{A}[\Omega] = -\frac{1}{16\pi G}\int_{t_1}^{t_2} \sqrt{-\bar{g}}\, d^4x\, [6\Omega^i\Omega_i - \bar{R}\Omega^2] + \mathscr{A}_m. \tag{3}$$

Here $\mathscr{A}_m$ denotes the action for the matter part. The study of quantum fluctuations in a singular spacetime, using (2) and (3) leads to the following general result: 'The quantum conformal fluctuations diverge at a classically singular event in a spacetime (Padmanabhan and Narlikar 1982). As discussed in I, this requires a revision of our classical concept of singularity.

Having thus tackled the first part of the analysis, it is necessary to introduce an alternative formalism for classical, geometrical description of singularities. This will be our concern in this paper.

What are the viable courses open for us? It is probably not out of place to discuss a similar dilemma faced by classical physics at the turn of this century. Classical physics predicted the collapse of an atomic electron into the nucleus in an exceedingly small

interval of time. This crisis was averted by quantum theory. Confining the electron too close to the atomic nucleus makes the momentum (and hence, energy) fluctuations diverge. Thus there arose the concept of a 'stationary state' for the atomic electron in which the quantum fluctuations stop the classical collapse. As a consequence there is a definite nonzero lowerbound on the electron's orbital radius.

This suggests the possibility of "stationary states for the spacetime geometry". Could it be that the quantum fluctuations lead to a lower bound in the scale of a collapsing universe? We shall show, in the sequel, that this is indeed true. The quantum structure of the collapsing universe is almost similar to the quantum structure of hydrogen atom!

2. Basic formalism

Classically, a particle is described by the trajectory $q(t)$; quantum mechanically, by a probability amplitude $\psi(q,t)$. In between these two extremes lies the 'correspondence principle limit' where the expectation values,

$$\langle q^n \rangle = \int dq \psi^*(q) q^n \psi(q), \tag{4}$$

can be used as a bridge between the two descriptions. The correspondence principle limit for our quantum gravity model is quite simple. It is characterised by the 'effective metric',

$$g_{ik}^{\text{eff}} = \langle \Omega^2 \rangle \bar{g}_{ik}. \tag{5}$$

Consider, for example, the closed Friedman model (discussed in I) with a background metric,

$$\overline{ds}^2 = dt^2 - \bar{S}^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\varphi^2)\right]. \tag{6}$$

The effect of quantum fluctuations is to replace the classical scale factor $\bar{S}(t)$ by,

$$S_{\text{eff}}^2(t) = \langle \Omega^2 \rangle S^2(t). \tag{7}$$

A direct computation shows that,

$$S_{\text{eff}}^2(t) \to L_p^2 \quad \text{as} \quad t \to 0. \tag{8}$$

(This is to be contrasted with the fact that $\bar{S}^2 \to 0$ as $t \to 0$, leading to the singularity.) Thus the quantum fluctuations indeed lead to a lower bound in the length scale.

It is the study of this example that motivated the concept of stationary states. The mathematical formalism is easy to arrive at, once we remember that, the stationary states exist only for time-independent Hamiltonians. Thus one has to treat the *total* conformal factor as the quantum variable in any static background metric. From the expression for action in (3) one can write down the Hamiltonian for the system, the eigen states of which, give the stationary states. We shall now present some examples of this formalism.

3. Examples of stationary states

We shall begin with the simplest case: conformal fluctuations about a vacuum background.

$$ds^2 = \langle \Omega^2(x) \rangle \, ds^2_{\text{flat}}. \tag{9}$$

The action for the conformal part has the simple form,

$$\mathscr{A} = -\frac{3}{8\pi G} \int (\Omega_{,i} \Omega^{,i}) \, d^4 x, \tag{10}$$

of a negative energy scalar field. This system can be quantised by the standard technique (Feynman and Hibbs 1965). Put

$$\Omega(\bar{x}, t) = \int a_{\mathbf{k}}(t) \exp(i\mathbf{k} \cdot \mathbf{x}) \frac{d^3 \mathbf{k}}{(2\pi)^3} \tag{11}$$

to reduce the action to the form,

$$\mathscr{A} = -\frac{3}{8\pi G} \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \int_{t_1}^{t_2} dt [|\dot{a}_{\mathbf{k}}|^2 - |\mathbf{k}|^2 |\mathbf{a}_{\mathbf{k}}|^2]. \tag{12}$$

This is the action for a set of harmonic oscillators each of which can be quantised by the standard techniques. Most important among the stationary states is the 'vacuum state' in which all the harmonic oscillators are in the ground state. It is straightforward to derive the wave functional, for the 'ground state of the gravitational field' as,

$$\psi[\Omega(\mathbf{x})] = N \exp\left\{ -\frac{3}{8\pi} \frac{1}{L_p^2} \int \int d^3\mathbf{x} d^3\mathbf{y} \frac{\nabla_x \Omega \cdot \nabla_y \Omega}{|\mathbf{x} - \mathbf{y}|^2} \right\}. \tag{13}$$

It is easily seen that violent fluctuations of geometry can exist over regions of the size $\sim L_p$. This vacuum functional gives the mathematical characterisation of the 'space-time foam' concept at small distances.

As a next example, consider a homogeneous isotropic space-time of the form,

$$ds^2 = \langle \Omega^2(t) \rangle \left[dt^2 - \frac{dr^2}{1 - r^2/a^2} - r^2 (d\theta^2 + \sin^2\theta \, d\varphi^2) \right]. \tag{14}$$

(because of symmetries Ω cannot depend on the space coordinates). The action takes the form,

$$\mathscr{A} = -\tfrac{1}{2} M \int_{t_1}^{t_2} dt [\dot{q}^2 - \omega^2 q^2] + \mathscr{A}_m, \tag{15}$$

where (with $c = 1$)

$$q = a\Omega, \quad \omega = c/a, \quad M = V c^2/Ga^2. \tag{16}$$

In these expressions V may be taken to be the proper volume of the background three space. Before proceeding further one has to decide the form of the source. Classically the universe is radiation-dominated near the singularity. In order to maintain a close parallel we shall assume $\mathscr{A}_m$ to arise from isotropic radiation. This has the advantage that, because of conformal invariance, $\mathscr{A}_m$ is independent of q and can be dropped.

The action in (15) leads to the simple classical solution,

$$\Omega(t) \sim \sin \omega t, \tag{17}$$

which correctly represent the classical radiation-filled universe. Quantum-

mechanically, we have the 'stationary state' equation,

$$\hat{H}\psi = -\frac{\hbar^2}{2M}\frac{\partial^2\psi}{\partial q^2} + \tfrac{1}{2}M\omega^2 q^2\psi = E\psi, \tag{18}$$

which, of course, represent a quantum oscillator of mass M and frequency ω. The stationary states are parametrized by the integer $n = 0, 1, 2, \ldots$. The quantum geometry in the nth stationary state is described by the metric,

$$ds^2 = L_p^2(n+\tfrac{1}{2})[d\eta^2 - d\chi^2 - \sin^2\chi\,(d\theta^2 + \sin^2\theta\, d\varphi^2)]. \tag{19}$$

The lowerbound at Planck length is obvious.

How can one connect up such a 'strange' quantum metric with the classical universe we know of? This strangeness can be removed by noticing the fact that quantum theory replaces the definite trajectory—*via* an expansion factor—by a probability distribution. For large enough n, we know from the study of harmonic oscillator wave functions that $|\psi_n(q)|^2$ has a classical profile.

There is another way of regaining the classical limit. One can take the universe to be in such a state that $\langle\Omega\rangle \sim \sin\omega t$ (the classical evolution). For harmonic oscillator there exists a set of states called "coherent states" which has this property. One can take the universe to'be in a state $\psi(q,t)$ with a probability density,

$$|\psi(q,t)|^2 = N\exp -\frac{a^2}{2L_p^2}(q - q_0\sin\omega t)^2. \tag{20}$$

The coherent states have the remarkable property that the dispersion does not spread with time ("non-spreading wave packets with classical mean"). The quantum geometric in this state is described by the metric,

$$ds^2 = (q_0^2\sin^2\eta + L_p^2)[d\eta^2 - d\chi^2 - \sin^2\chi(d\theta^2 + \sin^2\theta\, d\varphi^2)]. \tag{21}$$

Notice that when $\hbar \to 0$ this metric goes over to the standard classical solution. However, near the classical singularity ($\eta \sim 0$), the quantum effects lead to a lower bound $[q(\eta = 0) = L_p]$.

One might legitimately ask: what is the present quantum state of the universe? Is the universe in the stationary state or in a coherent state or is it in some other curious mixture?

At this stage it is impossible to provide a complete answer to such a question. In fact one has to make sure that the questions are meaningful. Quantum-mechanically the state of a system becomes well-defined only after a measurement has been made on the system. Thus one should attribute some meaning to the concept of "measurement of the state of the universe by an observer". At this preliminary level it is difficult to make any definite statement regarding this deep conceptual question.

To summarize, we have demonstrated the existence of definite stationary states for the conformal factor in the closed Friedman universe. These states give, in some sense, a partial "explanation" for the avoidance of the singularities.

Are these stationary states a special feature of the Friedman universe? The answer seems to be 'no'. All the homogenous relativistic cosmological models—the Bianchi types—exhibit these features we have discussed above (Padmanabhan 1982). In that sense one may consider the concept of 'stationary states for the geometry' to be sufficiently general.

4. Towards a more complete theory

In the previous section we have separated the space-time metric into a conformal part and a background part. The conformal part was quantised and it was shown that it leads to the correct classical limit. But what about the background metric? A complete theory of quantum gravity must determine both the conformal and the background parts without any external input.

In this section, we shall present a possible generalization of the previous formalism to take care of the background metric. This formalism seems to possess some interesting features peculiar to quantum gravity.

Let us begin by examining the classical theory. In the standard Einstein action,

$$\mathscr{A} = \frac{1}{16\pi G}\int R\sqrt{-g}\,\mathrm{d}^4x + \int \mathrm{d}^4x\sqrt{-g}\,\mathscr{L}_m. \tag{22}$$

Let us substitute $g_{ik} = \Omega^2(x)\bar{g}_{ik}$ and vary both and Ω & $\bar{g}_{ik}$ as independent quantum variables. This leads to the equations,

$$\langle\Omega^2\rangle(\bar{R}_{ik} - \tfrac{1}{2}g_{ik}\bar{R}) + 6t_{ik} + \frac{16\pi G}{\langle\Omega^2\rangle\sqrt{-\bar{g}}}\frac{\delta(L_m\sqrt{-g})}{\delta g^{ik}} = 0. \tag{23}$$

$$\Box\,\Omega + \tfrac{1}{6}\bar{R}\Omega = \frac{8\pi G}{3\Omega^3}\frac{1}{\sqrt{-\bar{g}}}\bar{g}^{ik}\frac{\delta(L_m\sqrt{-g})}{\delta g^{ik}}, \tag{24}$$

where $t_{ik} = -\partial_i\Omega\partial_k\Omega + \frac{1}{2}\bar{g}_{ik}(\partial_a\Omega\partial^a\Omega)$. It is trivial to show that these equations are identical to the standard Einstein equations.

Our previous discussions dealt with the quantum version of (24) (*i.e.* we dealt with the action in (22) with Ω as the quantum variable). It is clear that the background metric can be determined through (23) which we have not considered. Since functions of Ω appear in (23) we have to use the expectation values to obtain a c-number equation. Thus our 'hybrid' theory is based on the following rules:

(a) Treat Ω as a full quantum variable in the background metric $\bar{g}_{ik}$ *i.e.* replace (24) by the path integral expression for the kernel,

$$K[\Omega_2, \Omega_1] = \int \mathscr{D}\Omega \exp\frac{i}{\hbar}\mathscr{A}[\Omega]. \tag{25}$$

(b) Replace (23) with the following equation involving the expectation values of the quantum variable

$$\langle\Omega^2\rangle(\bar{R}_{ik} - \tfrac{1}{2}\bar{g}_{ik}\bar{R}) + 6\langle t_{ik}\rangle + \left\langle\frac{16\pi G}{\Omega^2\sqrt{-g}}\frac{\delta(L_m\sqrt{-g})}{\delta g^{ik}}\right\rangle = 0. \tag{26}$$

(c) Expectation values depend on the state chosen for the geometry. Make the choice when possible suitably to satisfy the above conditions.

It turns out that this version of the theory is much more restrictive than the classical gravity. Only certain forms of the metric and matter variables are allowed by (25) and (26).

For example, consider the isotropic homogenous (FRW) models discussed in the last section. Classically any amount of radiation will lead to a Robertson Walker universe

Quantum gravity does not allow this latitude. The background metric,

$$\overline{ds}^2 = \bar{g}_{ik}dx^i dx^k = \left[dt^2 - \frac{dr^2}{1-r^2/a^2} - r^2(d\theta^2 + \sin^2\theta d\varphi^2) \right], \tag{27}$$

with a source of isotropic radiation, (energy density ε)

$$T^i_k = \text{dia}\left(\varepsilon, \frac{-\varepsilon}{3}, \frac{-\varepsilon}{3}, \frac{-\varepsilon}{3}\right), \tag{28}$$

can satisfy (25) and (26) only if the following condition is satisfied:

$$\varepsilon = \frac{3}{8\pi^2}(\hbar c/a^4)(n+\tfrac{1}{2}). \tag{29}$$

In other words the energy density of the source automatically comes out to be quantised.

We know that Einstein's equations determine the classical dynamics of the source. Can quantum gravity follow this tradition and determine the quantum dynamics of the source? This attractive possibility remains to be explored.

It is clear from (29) that the equations dealt with are more restrictive than Einstein's equations. In fact, they are so restrictive that a flat vacuum without matter is not a solution to these equations, because of the $\langle t_{ik} \rangle$ term. In other words quantum gravity demands the existence of matter!

Contrast this situation with classical gravity. Given a matter distribution Einstein's equations determine the evolution of matter and the geometry. But a flat vacuum (with no matter) is a perfectly consistent solution to classical equations. Thus Einstein's equations are silent as to how the matter came into being. Mathematically, Einstein's theory guarantees the conservation of energy and thus is incapable of creating matter.

The situation is different as regards (26). Notice that $\langle t_{ik} \rangle$ has the form of a negative energy scalar field. Since only the contribution $(t_{ik} + T_{ik})$ is conserved, matter can be created or destroyed. This arises near the regions where the quantum fluctuations are important. Thus the above equations have in them an attractive method for creating matter near the singularities.

5. Conclusion

What we have presented is certainly not a complete theory of quantum gravity. However, even at this stage, one gets a glimpse of the rich structure of quantised gravity. Only further work will decide how tenable this approach is to quantum gravity.

References

Feynman R P and Hibbs A R 1965 *Quantum mechanics and path integral* (New York: McGraw-Hill)
Hawking S W and Ellis F R 1973 *The large scale structure of spacetime* (Cambridge: University Press)
Padmanabhan T 1982 *Gen. Rel. Grav.* (to be published)
Padmanabhan T and Narlikar J V 1982 *Nature* (*London*) **299** 675

Discussion

V. M. Rawal: What is the nature of fluctuations you consider?

T. Padmanabhan: The fluctuations we talk about are purely quantum mechanical. They may be taken to be the gravitational analogue of the vacuum fluctuations of, say, electromagnetic field.

D. R. Roy: The stationary states are Einstein's universes. Does it imply that the cosmological constant needs to be considered?

T. Padmanabhan: No. The solution is purely quantum mechanical. When the classical limit is taken no cosmological constant term survices.

J. R. Rao: Is $\langle \Omega \rangle$ a scalar?

T. Padmanabhan: Yes.

J. R. Rao: Is the homogenity of space connected with the stationary states considered in this model?

T. Padmanabhan: Stationary states can be defined for a wide variety of models. It is related to the existence of a static background but not directly to the homogeneity.

ISBN 9971-966-67-